Improved galactic foreground removal for B-mode detection with clustering methods

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **511,** 2052–2074 (2022) https://doi.org/10.1093/mnras/stac069
Advance Access publication 2022 January 12

Improved galactic foreground removal for B-mode detection with clustering methods

Giuseppe Puglisi [1,2,3]★ Gueorgui Mihaylov,[4] Georgia V. Panopoulou [5]† Davide Poletti,[6,7] Josquin Errard,[8] Paola A. Puglisi[9] and Giacomo Vianello[10]

[1]*Computational Cosmology Center, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[2]*Space Sciences Laboratory at University of California, 7 Gauss Way, Berkeley, CA 94720, USA*
[3]*Department of Physics, University of California, Berkeley, CA 94720, USA*
[4]*GSK and King's College London, Department of Mathematics, Strand, London WC2N 5RJ, UK*
[5]*California Institute of Technology, MC350-17, 1200 East California Boulevard, Pasadena, CA 91125, USA*
[6]*Universitá degli Studi di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[7]*INFN – Sezione di Milano Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[8]*AstroParticule et Cosmologie, Univ. Paris Diderot, CNRS/IN2P3, CEA/Irfu, Obs de Paris, Sorbonne Paris Cité 75005, France*
[9]*Royal Mail, 185 Farringdon Rd, London EC1A 1AA, UK*
[10]*Cape Analytics, 100 W Evelyn Ave UNIT 220, Mountain View, CA 94041, USA*

Accepted 2022 January 4. Received 2021 December 31; in original form 2021 October 4

ABSTRACT

Characterizing the sub-mm Galactic emission has become increasingly critical especially in identifying and removing its polarized contribution from the one emitted by the cosmic microwave background (CMB). In this work, we present a parametric foreground removal performed on to sub-patches identified in the celestial sphere by means of spectral clustering. Our approach takes into account efficiently both the geometrical affinity and the similarity induced by the measurements and the accompanying errors. The optimal partition is then used to parametrically separate the Galactic emission encoding thermal dust and synchrotron from the CMB one applied on two nominal observations of forthcoming experiments from the ground and from the space. Moreover, the clustering is performed on tracers that are different from the data used for component separation, e.g. the spectral index maps of dust and synchrotron. Performing the parametric fit singularly on each of the clustering derived regions results in an overall improvement: both controlling the bias and the uncertainties in the CMB B-mode recovered maps. We finally apply this technique using the map of the number of clouds along the line of sight, $\mathcal{N}_c$, as estimated from H I emission data and perform parametric fitting on to patches derived by clustering on this map. We show that adopting the $\mathcal{N}_c$ map as a tracer for the patches related to the thermal dust emission, results in reducing the B-mode residuals post-component separation. The code is made publicly available https://github.com/giuspugl/fgcluster.

Key words: cosmic background radiation – cosmological parameters – diffuse radiation – inflation – ISM – ISM: clouds.

1 INTRODUCTION

In the past few decades, the anisotropies of the cosmic microwave background (CMB) have been measured to constrain the parameters of the standard cosmological model, the Lambda cold dark matter (ΛCDM), with unprecedented precision (Planck Collaboration XIII 2016d; Planck Collaboration VI 2020b). In recent years, the polarization of the CMB has received significant attention. The linear polarization of the CMB arises primarily from Thompson scattering of photons with free electrons in the photon-baryon plasma of the epoch of recombination. The linearly polarized emission anisotropies can be decomposed into a field invariant under parity transformation, commonly referred to as the E modes, and a field that is not invariant known as B modes (Hu & White 1997; Seljak & Zaldarriaga 1997). The E modes are linked to the primordial scalar perturbations, whereas B modes at the degree scale can be produced only by tensor perturbations of the space–time metric emitted at the *inflationary* era (Guth 1981; Starobinsky 1982). B modes are expected to be observed at degree angular scales, making them an interesting scientific target to validate several theoretical models describing the early universe as their amplitude, commonly parametrized by the tensor-to-scalar ratio r, is proportional to the energy scale when inflation occurred.

At smaller angular scales (∼arcmin), B modes can be sourced by the direct distortion of E modes due to gravitational lensing of the intervening large-scale structure. Lensing B modes have been detected in the past years by many ground-based experiments (e.g. The Polarbear Collaboration et al. 2017, 2019; Choi et al. 2020; Bianchini et al. 2020). On the other hand, primordial B modes have not been detected yet and the best constraints have recently been set to $r < 0.044$, 95 per cent confidence level by Tristram et al. (2021), combining data from the BICEP2/Keck Array and Planck experiments.

★ E-mail: giuseppe.puglisi@uniroma2.it
† Hubble Fellow

Together with the instrumental sensitivity, the major challenge to detect primordial B modes is the polarized foreground radiation emitted at the same CMB frequencies ($\nu \sim 100\,\mathrm{GHz}$) from our own Galaxy. Electrons accelerated by spiralling along the Galactic magnetic field lines emit synchrotron emission mostly dominating the foreground polarized emission at low frequencies, $\nu < 70\,\mathrm{GHz}$. On the other hand, thermal dust emission arising from grains aligned with the Galactic magnetic field dominates the high frequency end of the sub-mm foreground radiation, $\nu > 100\,\mathrm{GHz}$. Polarization from other processes (e.g. molecular lines from Puglisi, Fabbian & Baccigalupi 2017) is expected to be subdominant with respect to that of Galactic synchrotron and dust emission (Planck Collaboration X 2016b; Planck Collaboration IV 2020a).

Since the Galactic components have different frequency dependences, they can be disentangled by means of component separation or foreground cleaning techniques (Planck Collaboration XII 2014; Planck Collaboration IX 2016a; Planck Collaboration X 2016c). On one hand, *blind* algorithms (e.g. internal linear combination, ILC) are usually aimed at removing emission that is different than the CMB, exploiting either the assumption of statistical independence for most of the sky components (see Delabrouille, Cardoso & Patanchon 2003; Maino et al. 2007; Remazeilles et al. 2018) or the maximum entropy principle (Stolyarov et al. 2005). On the other hand, *non-blind* approaches e.g. internal template subtraction or parametric fitting (Bennett et al. 1992; Hansen et al. 2006; Bobin et al. 2008; Eriksen et al. 2008; Leach et al. 2008; Stompor et al. 2008; Dunkley et al. 2009) are employed in order to distinguish several Galactic components. These come at the cost of making assumptions to model the frequency scaling of all the emission involved, commonly accounting for a very large number of free parameters. Moreover, what accentuates the difficulties of foreground removal is the fact that the properties of the interstellar medium (ISM) change not only *spatially*, i.e. in different locations in the sky (Planck Collaboration X 2016b; Planck Collaboration IV 2020a) but also along the same line of sight (Tassis & Pavlidou 2015; Clark 2018; Panopoulou & Lenz 2020), eventually leading to frequency dependencies that are more complex than typically assumed (Chluba, Hill & Abitbol 2017; Azzoni et al. 2021; Mangilli et al. 2021; Pelgrims et al. 2021). As a result, algorithms relying on parametric component separation would tend to reconstruct CMB maps with large residual foreground bias as the spatial variability of foregrounds is mis-modelled by fitting a single parameter across the whole observed sky.

A partition of the sky, where the parameters are fit independently on multiple regions, results in a mitigation of the bias but might increase the statistical uncertainties due to instrumental noise. This is mainly due to the fact that the fit is performed on a smaller number of pixels, encoding a lower signal-to-noise ratio (SNR) than the case with all the observed pixels. Moreover, performing a pixel-by-pixel foreground removal is clearly unfeasible for experiments involving large sky footprints and/or high resolution because fitting for spectral indices in each sky pixel is a very costly process.

Recently, several techniques to encompass the spatial variability of the foregrounds have been proposed by decomposing the map into wavelets or needlets (Basak & Delabrouille 2011; Remazeilles et al. 2018; Wagner-Carena et al. 2019; Irfan et al. 2019), by partitioning the sky into regions (clusters) according to the similarity of foreground properties and their location in the sky (Khatri 2019; Grumitt, Jew & Dickinson 2020).

In this work, we do spectral clustering on a proxy of SED properties of the Galactic foregrounds by considering the spectral parameters and the number of clouds along the line of sight map. Once the clusters are defined used them for component-separating multifrequency observations (Stompor et al. 2008; Errard, Stivoli & Stompor 2011; Stompor, Errard & Poletti 2016a; Alonso et al. 2017).

The approach is similar to the one shown in Grumitt, Jew & Dickinson (2020), but we choose a different *clustering method* to optimally divide the sky. In fact, Grumitt et al. (2020) used the *mean-shift* algorithm (Krzanowski & Lai 1971) to locate overdensities in the context of image segmentation. They developed a parametric bayesian component separation algorithm that makes use of a clustering analysis to forecast the accuracy of component separation for a LiteBIRD-like CMB space satellite mission (Sugai et al. 2020). They define clusters in a five-dimensional parameter space: three-Cartesian coordinates of pixels of the sphere, and the spectral indices β_s and β_d used to parametrize, respectively, the emission of synchrotron and dust. However, as noted by the authors, more work is needed to identify the optimal sky templates for this kind of study. An obvious area of improvement is to optimize the definition of sky regions. The characteristics of clusters obtained in Grumitt et al. (2020) were almost homogeneous across the sky: showing approximately the same size and similar (convex) shapes. However, the mean-shift cluster size is highly dependent on the choice of the bandwidth parameter in the definition of the kernels to identify the clusters. Ideally, the clustering should reflect the underlying spatial distribution of the features of interest. More sophisticated algorithms for image segmentation are available for this. In this work, we focus on improving upon the definition of clusters on the sky.

We propose in this work a different algorithm to partition the sky via *spectral clustering* (Von Luxburg 2007). In this specific application, the foreground spectral parameter maps are defined over the whole celestial sphere and can thus be mathematically treated as real-valued functions defined on the real manifold S^2. In the context of unsupervised image segmentation, spectral clustering has proved to be a powerful technique able to capture objects with highly non-trivial geometric shapes (Zhang et al. 2018; Zelnik-Manor & Perona 2004). Our approach takes into account the degree of similarity inferred both from geometric positions and from the measurements performed in different points of the sky. Being based on an eigen-decomposition, it is less affected by high-dimensionality issues. Furthermore, we use the signal-to-noise content of the spectral parameter maps to allow the pixel similarity to be informed by the intrinsic variability of a given foreground parameter map.

The definition of clusters on a manifold can be divided into two procedures: (i) building the pixel similarity accounting for the 'distances' in a given metric, (ii) finding the eigenspectrum decomposition for a given set of features. In Section 2, we present the spectral clustering methodology and the formalism to implement it on S^2. In Section 3, we present our implementation of spectral clustering that makes use of the HEALPix sky tesselation. In Section 4, we show two applications of how the patches defined with spectral clustering can be used to improve the performance of parametric component separation techniques. We finally discuss results and cosmological implications in Sections 5 and 6.

2 SPECTRAL IMAGE SEGMENTATION

Image segmentation via spectral clustering is an efficient technique that combines spatial similarity (or affinity) of pixels with the similarity based on image-related characteristics of the pixels (Zelnik-Manor & Perona 2004; Zhang et al. 2018). There are profound geometric and analytical reasons for this and are concisely discussed in the dedicated Appendix A.

Any image can be mathematically encoded by means of a real-valued (ideally smooth) function f, which is defined on a portion of a

surface and sampled with a fixed resolution given by the pixel grid. For example a grey-scale image is encoded by one function, whereas RGB images are encoded by three.

We summarize below the standard steps for a spectral clustering process:

(i) Translate an image into a graph by defining a suitable *adjacency* condition. A typical choice is that any pixel is connected to its nearest neighbours. For example, in correspondence to pixel j neighbour of i, we can set to a non-zero value the ij element of the so-called *adjacency* matrix A_{ij}. Once the connections among all pixels are defined via the weights on the adjacency matrix, we can build a graph by choosing a conventional order of the vertices (pixels) x_i. The way the adjacency weights are assigned is usually based on accounting for the geometric distance or the distance between the values that f takes in each pixel. An example of weights estimated from a distance is commonly weighting the graph nodes by a Gaussian function of the distance between two pixels.

(ii) Compute the graph Laplacian matrix from the weighted adjacency matrix. We use the symmetric random walk Laplacian:

$$L_{\rm sym} \equiv I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}, \tag{1}$$

where the I is the identity matrix and $D_i = \sum_j A_{ij}$ is a diagonal matrix called *degree* matrix.

(iii) Compute the eigenvectors of the graph Laplacian, $L_{\rm sym}$, and select a subset of n eigenvectors, in correspondence to the ones that contribute the most to the eigenspectrum of $L_{\rm sym}$. For example, for the specific case of 2D images, each eigenvector can be visualized as images encoding large (small) spatial variations depending whether the correspondent eigenvalue is small (large). This process defines an *embedding* of the graph associated with the image in $\mathbb{R}^n$, i.e. each pixel i in the image is characterized by n features, i.e. the value of each eigenvector in the ith pixel.

(iv) Once the embedding in $\mathbb{R}^n$ is defined, the standard Euclidean distance in $\mathbb{R}^n$ is evaluated for all the pair of pixels in order to run a suitable *Agglomerative* or *Divisive* clustering algorithm.

In this work, we combine the spectral clustering and adapted it in the context of quantities defined in the S^2 manifold. This is generally referred to as *manifold learning*: a discipline that combines statistical methods with techniques developed in differential geometry. It is based on the assumption that point clouds of multivariate n-dimensional variables are sampled on or close to smooth compact sub-manifolds of $\mathbb{R}^n$ (the existence of a boundary is usually ignored). In Appendix A, we outline more in detail how manifold learning can be combined with spectral clustering. However, we remark here two relevant and non-standard aspects of the implementation of the method proposed in this paper that make the geometric nature of spectral clustering very relevant:

(i) Maps from very wide astronomical surveys are sampled on the sphere, we therefore implement the affinity between pixels accounting for the angular distance on the S^2 sphere (presented in Section 3.1).

(ii) We choose the partition as the one that minimizes the within- and between-cluster variances. We present in Section 3.2, how the variance depends on several partition choices and eigenvector bands.

In order to better understand the details in the following sections, we briefly summarize few propositions in the context of Graph Theory which will be used to justify the choice for the adjacency in Section 3.

PROPOSITION 1. *The construction of the graph Laplacian matrix* L *is known to approximate the Laplace–Beltrami operator* $\Delta_{\rm g}$ *on a Riemannian manifold embedded in* $\mathbb{R}^n$ *and sampled in a grid of points (e.g. our set of pixels). Therefore, the weighted adjacency matrices can be directly derived from the integral kernel (or Green function) of the Laplacian operator in the sense of Fredholm's theory (Fredholm 1903).*

PROPOSITION 2. *The Laplace–Beltrami operator on a compact differentiable manifold* M *has discrete spectrum* $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{\rm n}$ *and its eigenfunctions* $\varphi_{\rm i}$ *form an orthonormal basis of the space* $\mathrm{L}^2(\mathrm{M})$. *Analogously, every discrete approximation of a function on a manifold can be expressed with respect to a basis of eigenvectors of the graph Laplacian.*

As a consequence of Proposition 1 and 2, estimating the eigen-spectrum of suitable graph Laplacian matrices sampled in a grid approximates the spectrum (eigenvalues and eigenfunctions) of the Laplace–Beltrami operator.

Moreover, in defining the optimal image segmentation it is quite common to define the *Dirichlet energy functional*, i.e.

$$E[f] \equiv \frac{1}{2} \int_M \parallel \nabla f(x) \parallel^2 \, {\rm d}\omega_g$$

that measures the spatial variability (or *spatial frequency*) of a smooth function on a manifold M. Thus, since the Dirichlet energy associated to the eigenfunctions of the Laplace–Beltrami operator is $E(\varphi_i) = \frac{1}{2}\lambda_i$ with monotonically increasing Dirichlet energy[1] (Proposition 2), the set of the first n eigenfunctions provides an optimal embedding of M in $\mathbb{R}^n$ corresponding to the minimum Dirichlet energy.[2] Generally, this process can be also addressed by selecting a subset of eigenfunctions (corresponding to a certain *Dirichlet energy band*) to embed M in a n-dimensional space.

The choice of the eigenfunctions related to a specific Dirichlet energy band can impact the quality of the image segmentation. This is mainly due to the fact that the clustering methodology is based on the choice of the eigenvectors of the Laplacian matrix, to which corresponds a certain degree of spatial variability, or *granularity*. We devote Section 3.2 to describe how this applies to the case presented in this work and how the choice of the energy band affects the characteristic size of the clusters.

The key ingredient in this framework is the definition of the adjacency, from which the Laplacian matrix can be then derived via equation (1). In the next section, we describe extensively the adjacency adopted in this work.

3 SPECTRAL CLUSTERING ON HEALPIX MAPS

In this section, we present an implementation of spectral clustering applied on images defined on the full celestial sphere. We adopt maps following the HEALPIX scheme (Górski et al. 2005; Zonca et al. 2019).

[1] A common example of increasing Dirichlet energy eigenfunctions can be found among the eigenfunctions of the Laplace–Beltrami operator in the sphere, i.e. the Spherical Harmonics, $Y_{\ell m}$. We observe increasingly larger spatial variability of the $Y_{\ell m}$ in correspondence to larger values of multipole number ℓ, with fluctuations involving smaller and smaller angular scales.
[2] Similarly, this properties propagates to the eigenvectors of the Laplacian matrix (equation 1).

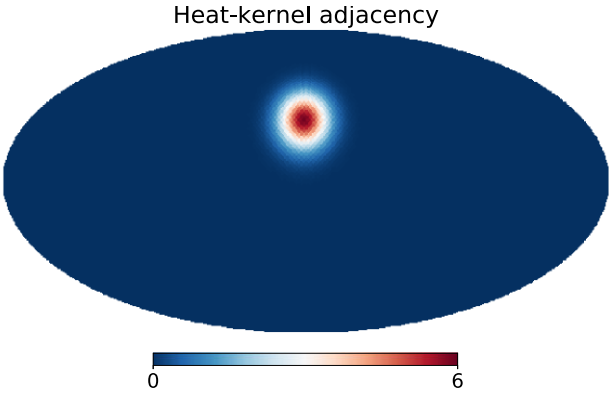

Figure 1. Adjacency as estimated in equation (A16) relative to the pixel centred at $(\ell, b) = (0, 30)$ deg Galactic coordinates. We recall the reader that A_{ij} is a $n_{\rm pix} \times n_{\rm pix}$, so its rows and/or columns can be visualized as HEALPIX maps.

3.1 Choice of the adjacency

Let us denote by $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_{\rm pix}})$, a set of normed-1 vectors (with $\boldsymbol{x}_i \in \mathbb{R}^3$), encoding the coordinates of each HEALPIX pixel. Given X, we can construct a matrix as $\Theta = X^T X$, being a $n_{\rm pix} \times n_{\rm pix}$ matrix, with diagonal equal to 1, and off-diagonal elements encoding the scalar product estimated pairwise on the columns of X. Notice that Θ encodes the cosines of the scalar product as the columns of X are normed-1.

We can therefore define the degree of affinity between pixels in a HEALPIX map from Θ, as

$$A \propto \exp\left(-(1-\Theta)^2/2\sigma^2_{\rm pix}\right), \tag{2}$$

where $\sigma_{\rm pix}$[3] is related to the pixel scale of the map. By expanding equation (2) in the Harmonic domain in terms of the Legendre Polynomials $\mathcal{P}_\ell$, as

$$A \equiv \sum_{\ell=0}^{+\infty} \frac{2\ell+1}{4\pi} \mathrm{e}^{-\ell(\ell+1)\sigma^2_{\rm pix}} \mathcal{P}_\ell(\Theta), \tag{3}$$

we observe that the scale given by $\sigma^2_{\rm pix}$ sets a threshold $\ell_{\rm max}$ to which we can stop the sum in equation (3).

In Appendix A4, we show how the functional form in equation (3) derives from the integral kernel (also known as *heat kernel*) of the Laplace–Beltrami operator, Δ_g, in S^2 Zhao & Song (2018). We show in Fig. 1 a row of the affinity matrix A as defined in equation (2).

When clustering methodologies are employed, it is common to ask oneself how the adjacency in equation (2) is affected in presence of the uncertainties in the measurements at a given pixel location. E.g. suppose that we have measurements of the *modified black-body* spectral index of dust emission, β_d, and its uncertainties, $\sigma_d \equiv \sigma(\beta_d)$. We require that two pixels encoding statistically compatible values of β_d would be more easily associated together into a single cluster with respect to pixels with incompatible values.

We therefore want to identify a way to combine the measurements and the uncertainties into the heat-kernel adjacency (equation 2) given realistic measurements of foreground spectral parameters. For the following application, we will use the map of the dust spectral index β_d and the uncertainty map σ_d obtained from the *Planck* data processed with the COMMANDER separation algorithm Planck Collaboration X (2016b).

Intuitively, we can distort Θ defined in equation (2) with a certain weight given by the measurements and the uncertainty of the parameter at a given pixel location. Given the state of art data on Galactic foregrounds in the sub-mm regime, we consider the measurements in each pixel i to be normally distributed around $\beta_{d,i}$ with a width given by its uncertainties, $\sigma_{d,i}$ (Planck Collaboration X 2016b). We therefore generate a mock sample of 100 Gaussian random numbers given the features $(\beta_{d,p}, \sigma_{d,p})$, and we estimate the two-sample Kolmogorov–Smirnov (KS) test on to pairs of different pixels i and j. This allows us to test the null hypothesis that the distribution of β^i_d values is drawn from the same underlying distribution as those of β^j_d within an assumed confidence level. We repeat the test 100 times, evaluate the statistical quantile Q_{ij} of each KS test, and set the median value all the KS tests, $Q^{\rm med}_{ij}$ as the adjacency weight evaluated for the ij pixel pair. Thus, if $Q^{\rm med}_{ij}$ is small enough to reject the null hypothesis at a 2σ confidence level, the connection is weighted with a very low weight and it is unlikely that the two pixels will be associated into the same cluster. On the other hand, when $Q^{\rm med}_{ij} \sim 1$ we cannot reject the null hypothesis and the two pixels can be associated, as they are *parallel* in this metric.

In order to combine the adjacency from the KS test and the one defined in equation (2), we can treat the KS quantiles as cosine of angles, so that a straightforward deformation of Θ can be

$$\Theta' = \cos\left(\arccos(\Theta) + \alpha\left(1 - Q^{\rm med}\right)\frac{\pi}{2}\right), \tag{4}$$

where $0 < \alpha \leq 0.5$ is a *scaling factor* parameter that weighs the relative contribution of the KS quantile similarity with respect to the geometrical one (the proximity of two pixels on the sky). Values of $\alpha > 0.5$ result into too large distortions that break the metric properties of equation (4) and make the Laplacian a singular matrix (see Section 3.2). The updated cosine matrix Θ' is then finally inserted into equation (2) to evaluate the adjacency weights obtained from this distortion.

Given the definition of adjacency (equation 2 or 4), we can therefore estimate the Laplacian matrix as in equation (1) and estimate its first $N_{\rm eigen}$ eigenpairs, related to the smallest eigenvalues. We remark that we do not factorize the whole Laplacian matrix, since we are interested in a subset of eigenvectors. We thus approximate the eigenpairs by means of the so-called *Ritz approximation*, a technique which has already been exploited in previous literature (Szydlarski et al. 2014; Puglisi et al. 2018) to approximate very well the exact eigenpairs of a matrix.

In Fig. 2, we show the first 100 Ritz eigenvalues λ of L for different choices of α. We firstly focus on the $\alpha = 0$ case. As already mentioned above, the functional form of the adjacency in equation (2) descends from the integral kernel of the Laplacian operator in S^2, which specifically coincides to the *angular momentum* operator, $\hat{L}^2$, in quantum mechanics (see Appendix A4 for further details). We thus expect the eigenvectors to be exactly the eigenfunctions of $\hat{L}^2$, the Spherical Harmonics, and to present the same algebraic multiplicity: for a given multipole number ℓ, its multiplicity goes as $2\ell + 1$, in correspondence of the m azimuthal multipole number.

Indeed, we notice that for the $\alpha = 0$ case (solid blue) line of Fig. 2, the algebraic multiplicity grows exactly with $2\ell + 1$. Moreover, the eigenvectors shown in Fig. 3 perfectly resemble the spherical harmonics. Both the eigenvalue degeneracy as well as the morphology of the Laplacian eigenvectors represent a remarkable validation test for the overall implementation described above.

[3] In this work, we adopted `nside=32` which results in $\sigma_{\rm pix} \sim 46'$.

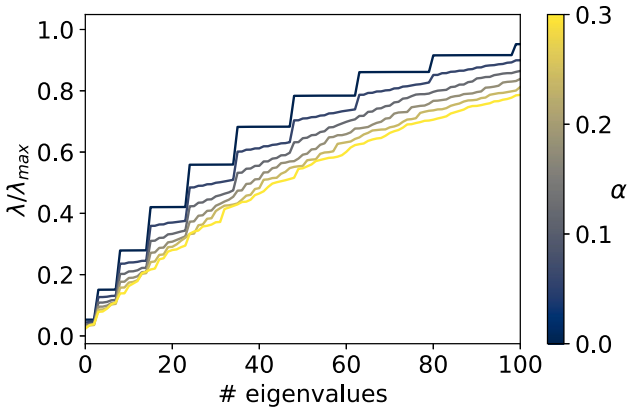

Figure 2. Eigenvalues of Laplacian matrix estimated as in equation (1), for several choices of the scaling factor α.

On the other hand, the distortion introduced by $\alpha \neq 0$ breaks the Laplacian eigenfunction algebraic multiplicity, so that the eigenspectrum increases more piecewise, for different higher values of α. As a consequence, the eigenvectors as well deviate from the spherical harmonics, being more and more weighted by the measurement uncertainties. In fact, for our specific case where Galactic emission is involved, we observe several anisotropies in correspondence of the Galactic plane (see Fig. 4).

Algorithmic 1. Spectral Clustering Optimization of a field X with uncertainty $\sigma(X)$

procedure SPECTRAL CLUSTERING ($X, \sigma(X)$)
  Parameter Affinity, Q Initialization with ($X, \sigma(X)$) data
  for α in $[0, \ldots, \alpha_{max}]$ **do**
    Build Affinity matrix with α distortion (eq. 4)
    Build Laplacian matrix, L (eq. 1)
    Estimate Ritz eigenpairs
    Select an eigenvector-band, U
    Estimate the euclidean affinity E from the columns of U
    for δ in $[0, \ldots, \delta_{max}]$ **do**
      Run Agglomerative clustering
      Estimate $W(\alpha, \delta)$
    end for
  end for
  Find local minima of W, (α_*, δ_*)
end procedure

3.2 Identifying the optimal partition

In this subsection, we aim at discussing how the clusters are estimated and how to identify the range of partition optimality. The overall spectral clustering algorithm described in the sections above is summarized in Algorithm 1. In particular, we employ the implementation of agglomerative clustering publicly available in the SCIKIT-LEARN PYTHON package[4] to estimate the clusters.

The relationship between the eigenvectors of the Laplacian and the spatial frequency of the associated eigenvectors (measured by the Dirichlet energy functional) is explained at the beginning of Appendix A. Lower eigenvalues correspond to eigenvectors with lower spatial variability (lower Dirichlet energy). The Dirichlet energy of the Laplacian eigenfunctions on a smooth manifold can increase indefinitely, while in the discrete case the Laplacian matrix reaches a maximum eigenvalue. The spectrum of the Laplacian matrix manifests concave shape for different values of the scaling factor α (see Fig. 2). The eigenvalues of the Laplacian matrix on S^2 increase at a decreasing rate while their algebraic multiplicity increases. Moreover, the algebraic multiplicity of the eigenvalues is broken by the geometric deformation introduced by α.

Since we want to identify the smallest eigenvalues encoding most of the bulk of the Laplacian eigenspectrum, we consider the first $n_{\rm eig} = 256$ Ritz eigenpairs and therefore build the matrix $U \in \mathbb{R}^{n_{pix} \times n_{eig}}$ containing $n_{\rm eig}$ eigenvectors as columns.

We remark here that the eigenvector matrix U is critical in the context of spectral clustering because it provides the embedding of S^2 into $\mathbb{R}^{n_{eig}}$ space spanned by the Laplacian eigenvectors. Moreover, it helps in reducing the dimensionality of our problem from $n_{\rm pix}$ features, i.e. the number of features we have in a map, to $n_{\rm eig}$ and in avoiding the *curse of high dimensionality* which clustering methodologies might frequently incur.

We then consider each row of U, $(u_i)_{i=1,\ldots,n_{pix}}$ as our data vector and we perform agglomerative clustering with $n_{\rm pix}$ samples and $n_{\rm eig}$ features. We measure the pairwise similarity, E_{ij}, between two vectors u_i and u_j (representing the ith and jth pixels) by estimating the euclidean distance between them. The two pixels are associated with a group if $E_{ij} < \delta$, with δ being a free parameter, commonly referred as the *linkage distance threshold*. Small values of δ tend to typically produce small clusters, and vice versa. This can be intuitively understood as follows: the more constraining the distance threshold, the harder it is to associate two pixels together into a cluster.

We notice that for several values of α the distance matrix range changes in such a way that we get smaller eigenvector similarities for larger values of α. This is a consequence of the fact that the deformation of adjacency with KS weights (equation 4) results into a compression of the Laplacian eigenspectrum to smaller condition numbers. In fact, we notice a trend in the Laplacian eigenspectrum to (i) lose the typical $2\ell + 1$ degeneracy for $\alpha \neq 0$, (ii) flatten to smaller condition number, (iii) to become degenerate to a small value numerically close to zero. As a results the convergence of the Laplacian Ritz eigenpairs estimation time increases for $\alpha > 0.3$. [5]

Furthermore, the eigenvector distance for several values of α needs to be normalized in such a way that we use the same range in δ to compare the variances estimated from each run of the agglomerative clustering. We therefore rescale the eigenvector distance E_α evaluated for different choices of α as it follows:

$$E'_\alpha = \frac{E_\alpha/M - m}{1 - m}, \tag{5}$$

with $M = E_{\max,\,\alpha}$ and $m = E_{\min,\,\alpha}$. We note that E' ranges from $E'_{\max} = 1$ to $E'_{\min} = Cm$, with $C = (1 - M)/M \sim$ constant for different values of α. The rescaling in equation (5) can be seen as a sort of *rigid* scaling of the distance, making the distance matrices for different choices of δ and α to be comparable without losing too much information. Notice that equation (5) scales the maximum of the distance to 1 but we make sure that it does not affect the shape of distance distribution. We also make sure that the ratio of minimum distances for different choices of α is the same as before the rescaling in equation (5).

[4] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html

[5] This is the reason why we set the maximum value of α to 0.3.

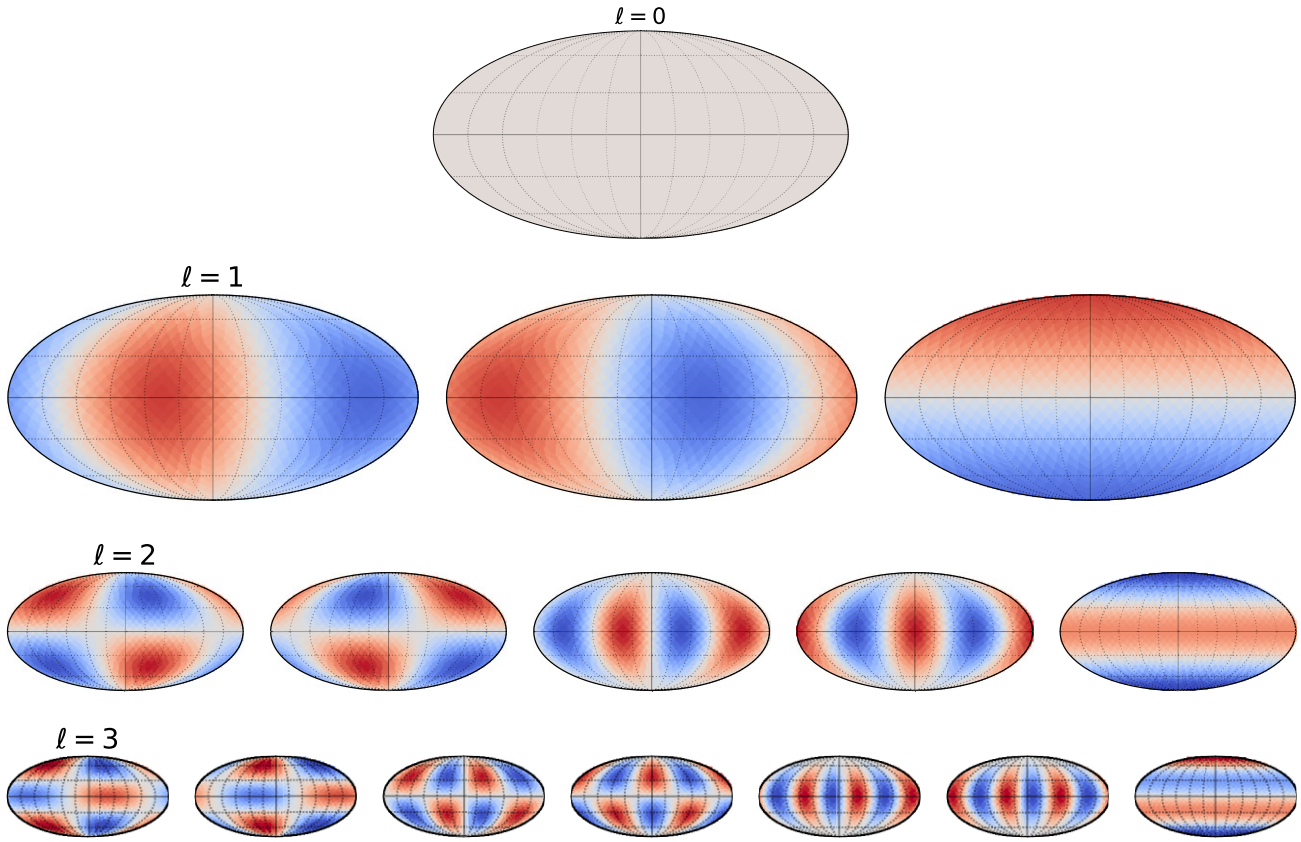

Figure 3. The first eigenvectors of L corresponding to $\ell = 0, \ldots, 3$ and assuming no uncertainty deformation (i.e. $\alpha = 0$). The colourbar ranges for all the maps between ± 0.07.

Furthermore, we consider three bands to identify the eigenvectors with major contribution to the total Dirichlet energy budget, i.e. the spatial variability, in the domain partition:

(i) ℓ_{b1}, including the first 100 eigenvectors, corresponding to $1 \lesssim \ell \lesssim 10$.
(ii) ℓ_{b2}, including vectors between the 90th and 150th eigenvectors, corresponding to $9 \lesssim \ell \lesssim 12$.
(iii) ℓ_{b3}, sampling the whole eigenspectrum every other 5 vectors up to the 175th one, corresponding to $1 \lesssim \ell \lesssim 14$.

We create a grid with different points evaluated in the $\alpha - \delta$ parameter space, with $\alpha \in [0, 0.5]$, $\delta \in [0.05, 0.3]$, and estimate the cluster variance, W, defined as

$$W \equiv \sqrt{V_w^2 + V_b^2}, \tag{6}$$

with V_w being the within-cluster variance,

$$V_w \equiv \sum_{k=1}^{K} \sum_{u_k \in C_k} (u_k - \mu_k)^2,$$

V_b the between cluster variance,

$$V_b \equiv \sum_{k \neq k'} (\mu_{k'} - \mu_k)^2,$$

K the total number of clusters and μ_k the centroid of cluster C_k.

Fig. 5 shows W defined in $\alpha - \delta$ space evaluated for the three different eigenvector bands. The optimal parameters (α_*, δ_*) correspond to local minima of W. Notice that the range of optimality changes for different eigenvector bands as different scales are encoded in each band. We emphasize that similar optimal ranges are found among the three bands, indicating that the clustering is stable with respect to the choice of the band.

Since the ℓ_{b3} band encodes scales from most of the Laplacian eigenspectrum, we adopt this band as a baseline for the applications presented in the following sections.

Finally, we notice that the optimality, shown in Fig. 5, does not correspond to a very localized region in the parameter space spanned by α and δ. Rather it shows as a narrow vertical band at around a certain value of δ for which several choices of α can be equally optimal. We find that intermediate values of $\alpha \sim 0.15$ lead to a balanced trade-off between defining the clustering, given the features of spectral parameters and the intrinsic adjacency in S^2. We refer the reader to Appendix C for further details.

The estimation of the Laplacian adjacency and eigenpairs and the optimization of spectral clustering has been performed on 300 Cori KNL nodes of the NERSC Supercomputing facility.[6] It takes ∼4000 cpu-hours to execute the overall procedure outlined in Algorithm 1 with maps at `nside=32`.

4 APPLICATIONS

We present two applications of our spectral clustering algorithm for the purpose of performing parametric component separation. First, in Section 4.1, we apply the spectral clustering algorithm to derive a partition of the sky from publicly available maps of synchrotron

[6] https://www.nersc.gov/

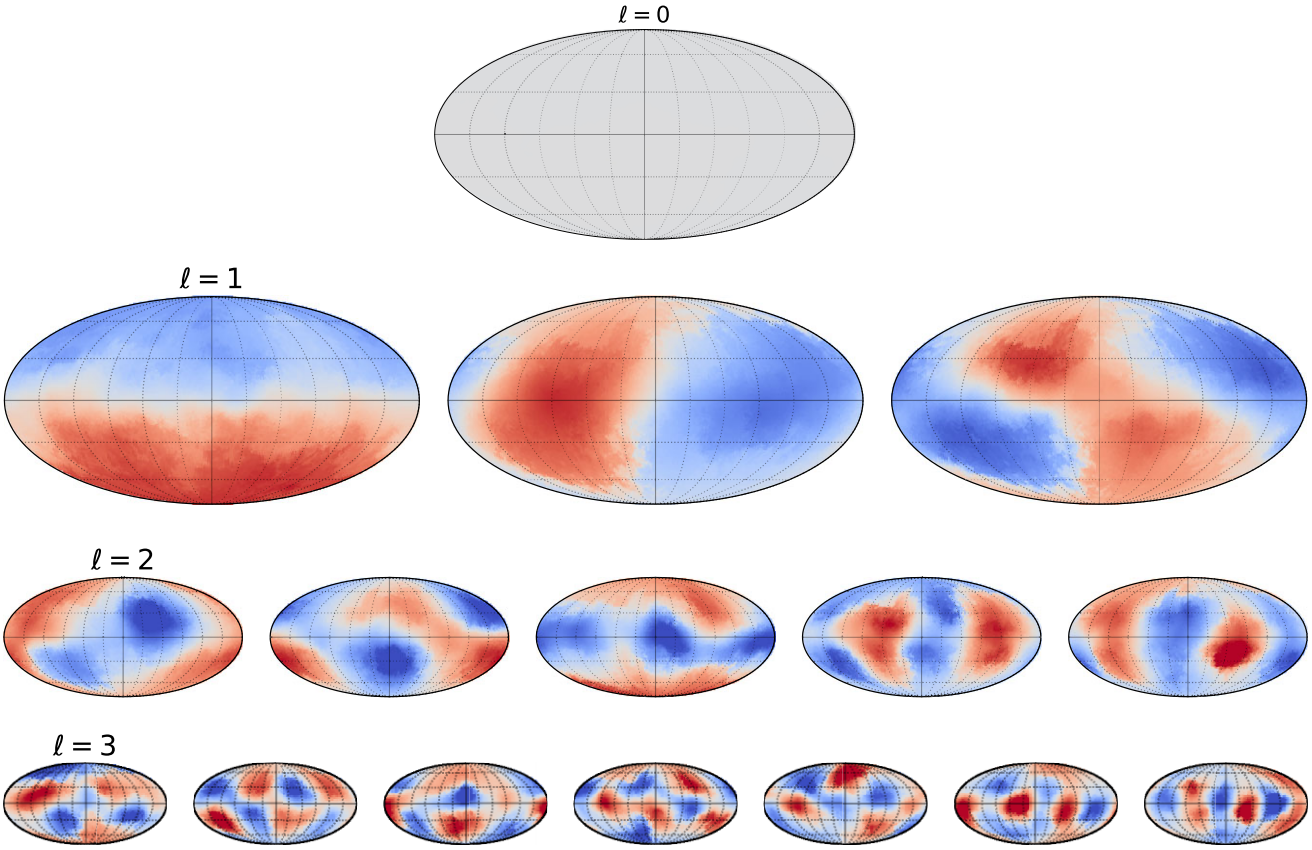

Figure 4. The first eigenvectors of L with affinity deformation $\alpha = 0.25$. For sake of comparison with Fig. 3, we report also in this plot the correspondence to $\ell = 0, \ldots, 3$ although the analogy lose specific meaning for $\alpha \neq 0$. The colorbar ranges for all the maps between ± 0.02.

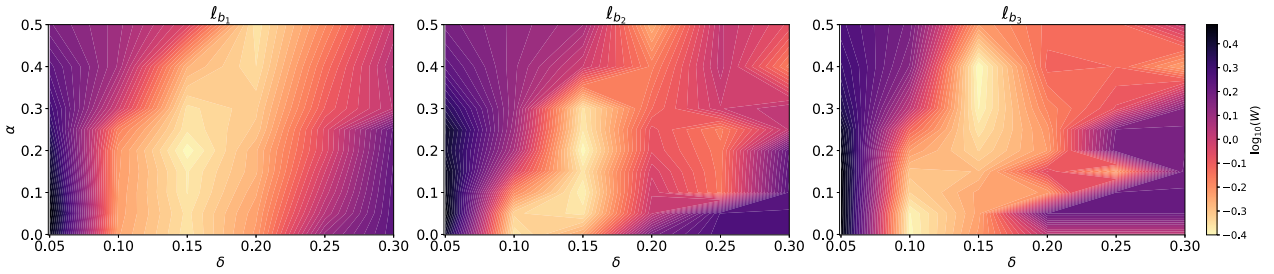

Figure 5. Variance W as in equation (6) in the $\alpha - \delta$ space, evaluated in clusters constructed with the three eigenvector bands as defined in the main text. For the analysis shown in Section 4, we adopted the ℓ_{b3} being the band that accounts for the eigenvectors encoding most of the angular scales to be clustered.

and dust spectral parameters. Once the patches are defined we perform a parametric component separation on each patch and assess the post-foreground cleaning residuals in the CMB B-mode maps. Secondly, in Section 4.2, we identify the patches by means of an ancillary data set: the number of H I clouds along the line of sight, $\mathcal{N}_c$, presented in Panopoulou & Lenz (2020). This data set is an independent tracer of the foreground dust emission. We then perform component separation within these patches and estimate the quality of the reconstruction.

4.1 Clustering applied on Galactic foreground spectral parameters

We perform parametric component separation by fitting for three components: CMB, dust, and synchrotron emission. We consider two different combinations of data, to be representative of the forthcoming experiments aimed at observing primordial B modes: the Simons Observatory Small Aperture Telescope (SO-SAT; Simons Observatory Collaboration et al. 2019) from the ground and the LiteBIRD space satellite (Sugai et al. 2020). The component separation is performed, respectively, on 27 per cent and 60 per cent of the sky (see Fig. 9) with frequency channels encoding the specifics shown in Table 1. Notice that for SO-SAT we assume conservatively the *baseline* configuration as described in Simons Observatory Collaboration et al. (2019).

We use the Python Sky Model[7] (PySM; Thorne et al. 2017) to simulate full-sky polarized emission of thermal dust, CMB, and

[7] https://pysm3.readthedocs.io/

Table 1. Nominal specifics of CMB experiments to run component separation with FGBUSTER.

	Frequency (GHz)	Sensitivity (μK arcmin)	FWHM (arcmin)
SO-SAT	27, 39, 93	35, 21, 2.6	91, 63, 30
	145, 220, 270	3.3, 6.3, 16	17, 11, 9
LiteBIRD	40, 50, 60, 68, 78	37.5, 24, 19.9, 16.2, 13.5	69,56,48,43,39
	89, 100, 119, 140, 166	11.7, 9.2, 7.6, 5.9, 6.5	35, 29, 25, 23, 21
	195, 235, 280, 337, 402	5.8,7.7,13.2,19.5,37.5	20, 19, 24, 20, 17

synchrotron.[8] The synchrotron radiation is commonly parametrized as a power law

$$I_{\nu,\mathrm{synch}} \propto \nu^{\beta_s}, \tag{7}$$

whereas the thermal dust emission is described by a modified black-body, i.e.

$$I_{\nu,\mathrm{dust}} \propto \nu^{\beta_d} B_\nu(T_d), \tag{8}$$

with T_d being the black-body dust temperature.

Among the different models available in PySM, we choose as a reference the `d1s1` model, which accounts for spatial variation in the synchrotron and dust spectral indices, β_s, β_d, and in T_d.

The synchrotron spectral index of the `s1` model is the Miville-Deschênes et al. (2008, model 5), estimated by combining the Haslam et al. (1982) map at 408 MHz and the map at 23 GHz from WMAP (Hinshaw et al. 2009). The β_s map and its uncertainties $\sigma(\beta_s)$ are available online[9] at 5 deg resolution.

The `d1` model is derived from the *Planck* data products[10] released in Planck Collaboration X (2016b), with β_d and T_d parameters obtained at about 1 deg resolution with the COMMANDER component separation algorithm Planck Collaboration X (2016b).

When multiple frequency channels are simulated with the `d1s1` model, algorithms relying on parametric component separation would tend to reconstruct CMB maps with large residual foreground bias (commonly referred as *systematic bias*) if a single constant spectral parameter is fitted for each Galactic component across the whole observed sky, as the spatial variability of foregrounds is not taken into account in the fit. Several approaches in the literature tried to tackle this bias, in the literature the most common one is Commander (Eriksen et al. 2006, 2008). Commander employs a Bayesian parameter fit to estimate the free parameters by means of Markov Chains with Gibbs sampler and non-linear searches. It has successfully been applied on *Planck* real data both in intensity and polarization. Although it provided the state-of-art maps of synchrotron and dust parameters with spatial variation at the $\sim$ degree scale, Commander can work only on coarse pixel grids, as its computational cost scales with the number of pixels. Lately, Khatri (2015, 2019) implemented a novel methodology combining clustering and ILC methods to intensity only *Planck* data to exactly estimate high-resolution spectral parameter maps.

On the other hand, the instrumental noise is responsible for the *statistical* uncertainties that can be commonly assessed by running component separation on frequency maps encoding different Monte Carlo (MC) realizations of instrumental noise and astrophysical signal.

A partition of the sky where the parameters are fit independently on multiple regions might mitigate the systematic bias but increase the statistical uncertainties, as the fit is then performed on a smaller number of pixels, hence encoding a lower signal-to-noise ratio (SNR) than the case with all the observed pixels (Errard & Stompor 2019).

Once the set of parameters is defined together with their associated uncertainties, we can run the spectral clustering optimization procedure as outlined in Section 3. The variance planes for β_s, β_d, T_d are shown in Fig. 6. We show the optimal patches selected in correspondence of local minimum of the variance in Fig. 7.

We estimate the median value of the number of pixels included in each cluster and we find about 4 pixels per cluster for β_d, 8 for T_d, and 16 for β_s, resulting in a typical angular sizes for the cluster range from $\sim$ 3.5, 7, 14 deg, respectively, for β_d, T_d, and β_s. The different morphologies and sizes, as well as the different number of clusters, result from a trade-off between the intrinsic variability due to the astrophysical emission, the resolution of each map (see the left column of Fig. 8) and the SNR. Specifically, the clustering algorithm finds it harder to agglomerate pixels related to very high SNR regions (e.g. at low Galactic latitudes) with respect to regions with larger uncertainties. In fact, this trend is observed in all the regions: smaller patches where the SNR is high, larger patches where it is low.

We then implement a new functionality in FGBUSTER[11] (Stompor, Errard & Poletti 2016b), a package for parametric component separation, aimed at fitting parameters within the set of multiple partitions of the sky. Further details of the implementation will be presented in two companion papers (Errard et al. in preparation; Poletti et al. in preparation).

We adopt the usual emission model assumed parametric fitting. For a given pixel p, each map at a given frequency, $m(\nu)$, can be expressed as

$$\begin{aligned} m^{p,X}(\nu) = {} & A_s^{p,X} f_s(\nu, \beta_s) + A_d^{p,X} f_d(\nu, \beta_d, T_d) \\ & + A_{\mathrm{cmb}}^{p,X} f_{\mathrm{cmb}}(\nu) + n^{p,X}(\nu), \end{aligned} \tag{9}$$

i.e. a linear combination of amplitudes of each astrophysical component, with functions $f_s(\nu, \beta_s)$, $f_d(\nu, \beta_d, T_d)$, $f_{\mathrm{cmb}}(\nu)$ encoding the spectral dependence of synchrotron, dust, and CMB, a at a given frequency ν as outlined in Section 4.1; $n(\nu)$ accounts for instrumental noise of a given frequency channel coadded to the signal. For a given free parameter, the fit is performed on different sky regions. For example, a single value of β_d is estimated within each cluster shown in Fig. 7 (left). Whilst, T_d value in the same location is fitted from a different set of pixels with respect to β_d and β_s (see for a comparison the patches in Fig. 7). This mainly implies that each spectral parameter in the same sky location is estimated by considering a different set of pixels from the multifrequency maps with respect to the ones employed for another parameter. Essentially, each fit performs *in an adaptive way* given the intrinsic morphology, size and SNR of the emission. This is the reason why we expect large improvements in terms of parametric evaluation.

[8] The contribution in polarization from AME, CO and free–free is expected to be very small and it is neglected in the following.
[9] https://lambda.gsfc.nasa.gov/
[10] https://pla.esac.esa.int
[11] https://fgbuster.github.io/fgbuster

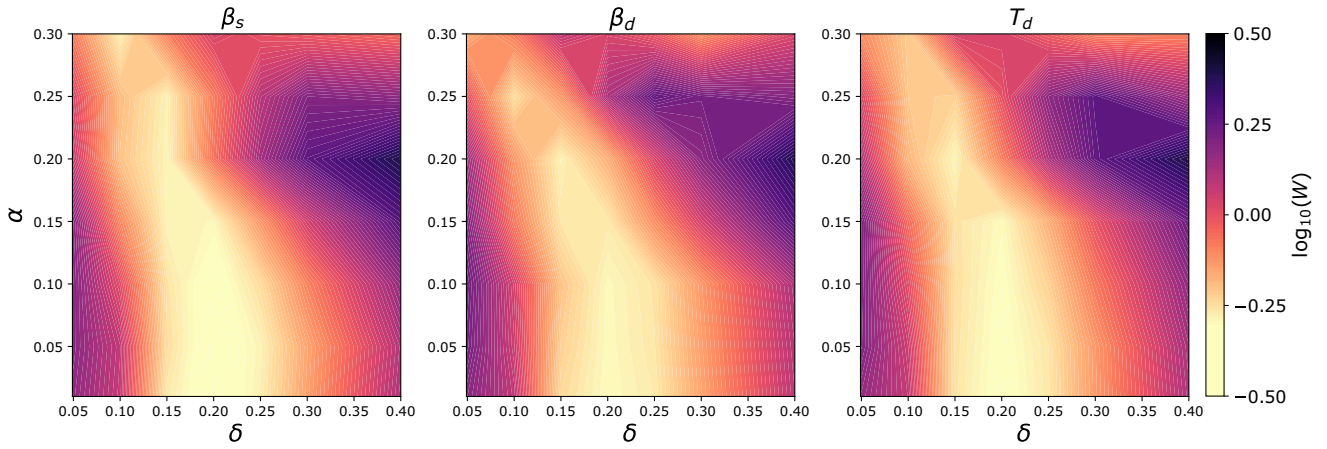

Figure 6. Variance surfaces estimated for several choices of α and δ, for (left) β_s, (middle) β_d, (right) T_d. Note that regions of optimality correspond to local minima (lighter regions) in the surfaces.

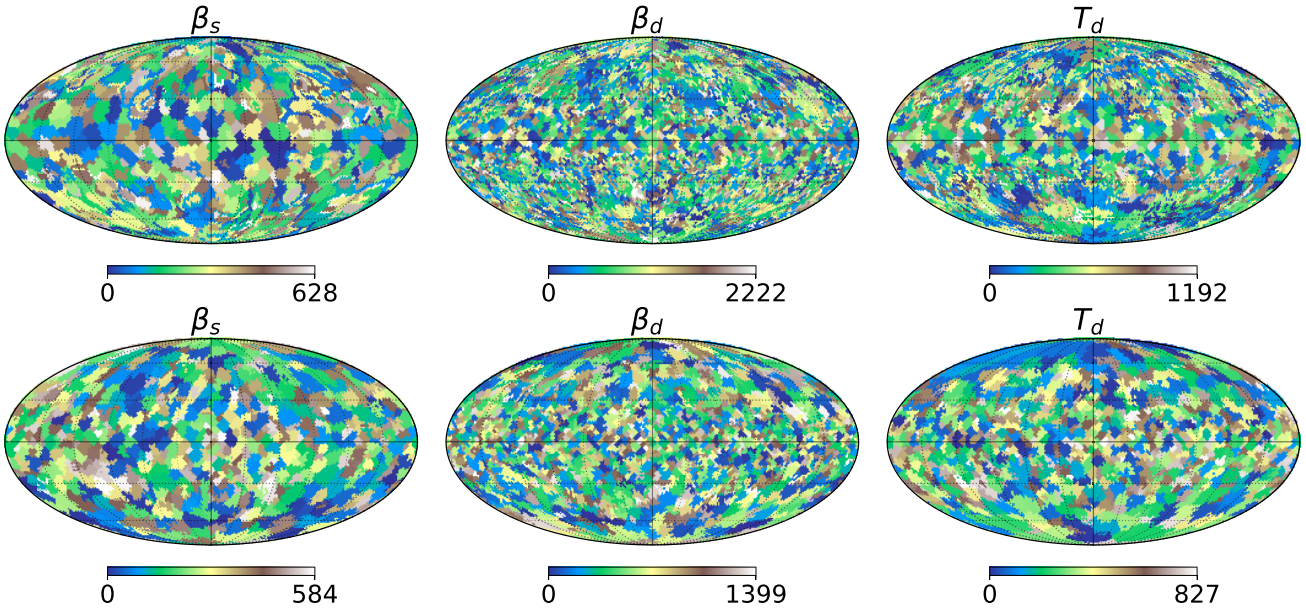

Figure 7. Cluster-defined patches obtained from the parameters in the `d1s1` model (each colour denotes pixels belonging to different clusters). We show Clusters estimated with $f = 0$ (top row) and with $f = 1$ (bottom row).

We generate 20 Monte Carlo realizations of noisy frequency maps with the nominal specifications of LiteBIRD and for SO-SAT (see details in Table 1). The frequency maps are produced with the PySM `d1s1` model at `nside=64` and include both the input astrophysical polarized signal and the instrumental noise (assumed to be only white noise). We show in Fig. 9 the nominal sky patches where the parametric fit is performed. Notice that we choose a larger $f_{\rm sky} = 0.27$ for SO-SAT with respect to the effective $f_{\rm sky} = 0.10$ reported in Simons Observatory Collaboration et al. (2019). The reason for this choice is mainly motivated by the fact that we want to conservatively assess the performance of this methodology on to larger regions of the sky being thus more susceptible to larger Galactic residuals. This choice further allows us to compare our results with the ones reported in Thorne et al. (2019).

In Fig. 8 (middle column), we show the parameter maps estimated in each cluster region with the LiteBIRD frequencies and we evaluate the relative error of this estimate by considering the difference between the PySM parameter maps and the parameters estimated with FGBUSTER (see the right column in Fig. 8).

In Fig. 10, we show the distributions of relative errors for the three parameters together with the median values shown as vertical dashed lines to be 0.002, 0.006, 0.017, respectively, for β_s, β_d, T_d, with the largest errors, ≤ 20 per cent, found in T_d estimates. This implies that the component separation performed on the cluster-defined patches in Fig. 7 (top) recovers faithfully the spatially varying Galactic components. Furthermore, we somewhat expect the worst estimate to be the one related to T_d, since both LiteBIRD and SO-SAT high-frequency channels probe frequency regimes far from the modified black-body peak ($\nu \sim 600$ GHz), making T_d essentially degenerate with β_d at lower frequencies.

The case shown in Fig. 8 is an ideal-limit case as the patches have been derived by clustring on to the same parameter maps that are used to simulate the frequency channels. In order to introduce a *mis-modelling* between the two set of maps, we inject a noise proportional

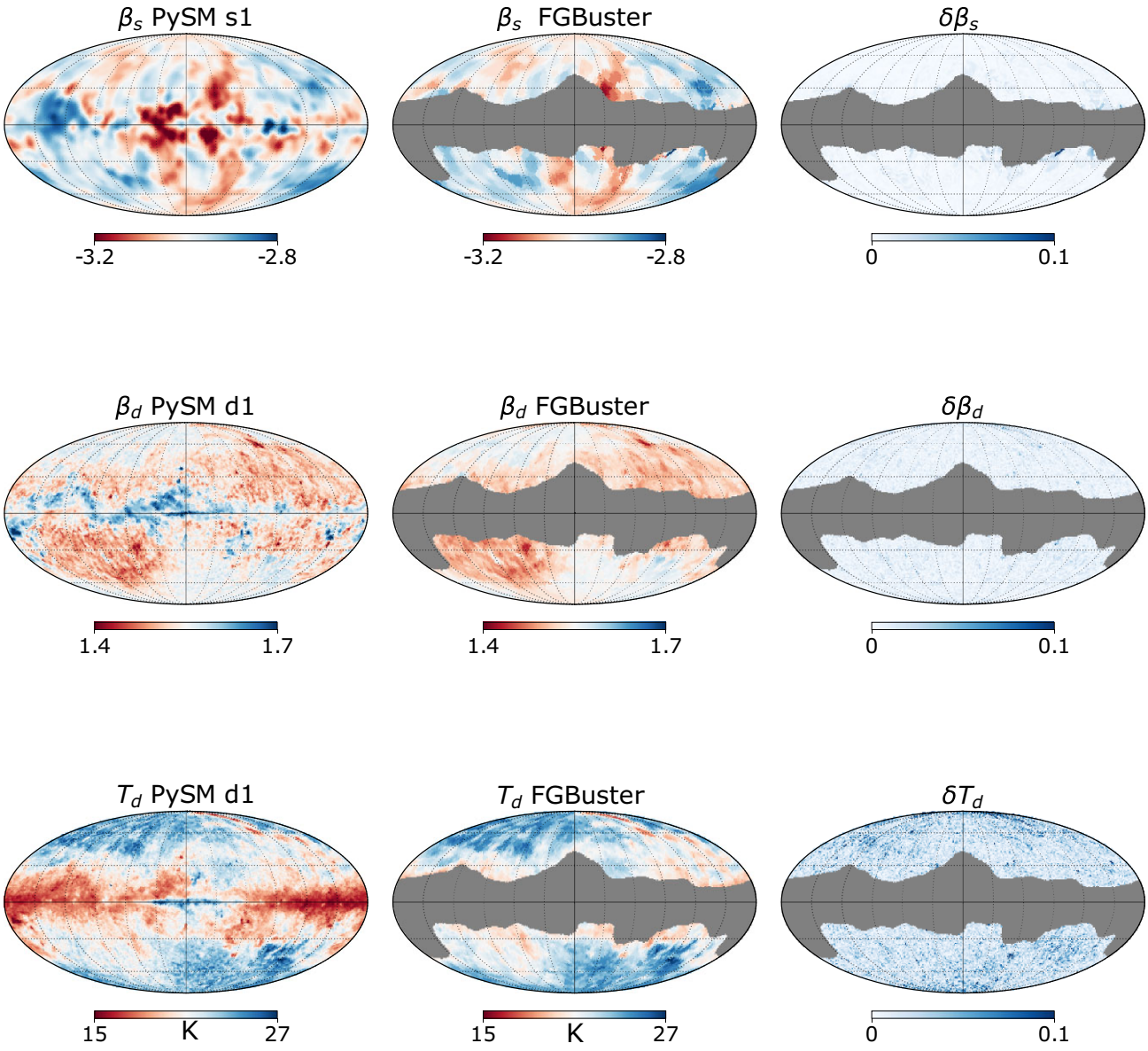

Figure 8. Left column: Foreground parameters from the PySM `d1s1` model, (middle column) parameter maps estimated with FGBuster on each clustering region (see Fig. 7) using the LiteBIRD frequency channels. Right column: Relative residuals on the estimated maps with FGBuster and the input ones from PySM.

to the per-pixel uncertainties on β_s, β_d, T_d, as

$$\tilde{X} = X + f\ N_w[\sigma(X)], \tag{10}$$

where $X = \beta_s$, β_d, T_d; $N_w[\sigma(X)]$ is a random Gaussian noise map with zero mean and width given by the uncertainty map, and f is a weighting constant factor, ranging from 0 to 1.

The $f = 1$ case represents a pessimistic case where the foregrounds are fully mis-modelled, given the uncertainty budget in the spectral parameters and $f = 0$ represents the ideal case without any mis-modelling.

For each mis-modelling case, we estimate the statistical residuals by performing component separation on maps encoding 20 MC independent realizations of instrumental noise and astrophysical signal. The systematic bias is instead estimated by running component separation on noiseless maps.

We therefore run the spectral clustering algorithm on the mis-modelled parameter maps $\tilde{\beta}_s$, $\tilde{\beta}_d$, $\tilde{T}_d$ for several values of f. For example, in Fig. 7 (bottom), we show the cluster obtained from parameter maps with maximum mis-modelling ($f = 1$). We note a trend to have smaller number of clusters with similar sizes and more uniformly distributed with respect to the $f = 0$ case for similar values of α, δ. This is particularly noticeable for the dust parameters, since they are provided with a higher resolution than the synchrotron one. We conclude that injecting the mis-modelling noise into the spectral parameter maps tends to homogenize the typical size of the patches.

We then perform the component separation separately for LiteBIRD and SO-SAT channels (simulated without mis-modelling) on to the cluster patches derived with the mis-modelling. This procedure allows us to further set requirements on the mis-modelling given the residual bias we observe in the recovered CMB map for different values of $f = [0, 0.1, 0.5, 1]$.

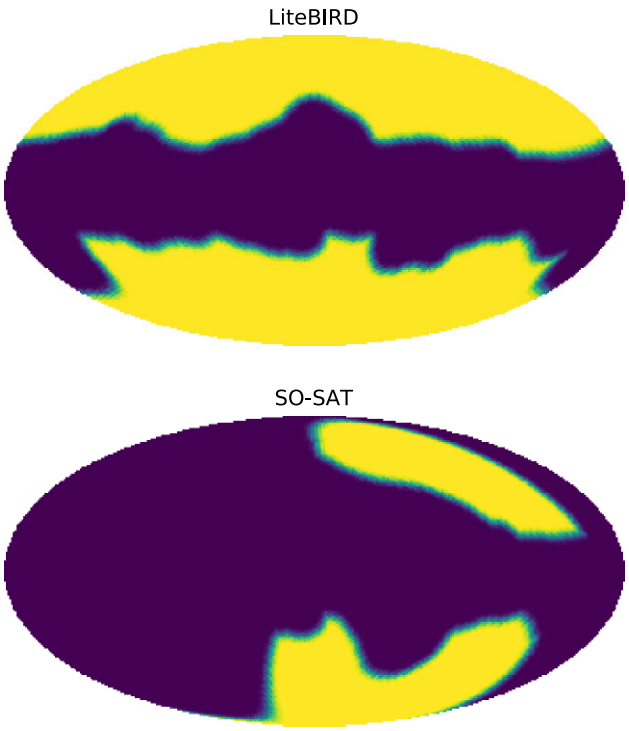

Figure 9. Masks used for component separation and for the power spectra estimation with NaMaster (Alonso, Sanchez & Slosar 2019) for LiteBIRD (top) and SO-SAT (bottom) encoding, respectively, 60 per cent and 27 per cent of the sky.

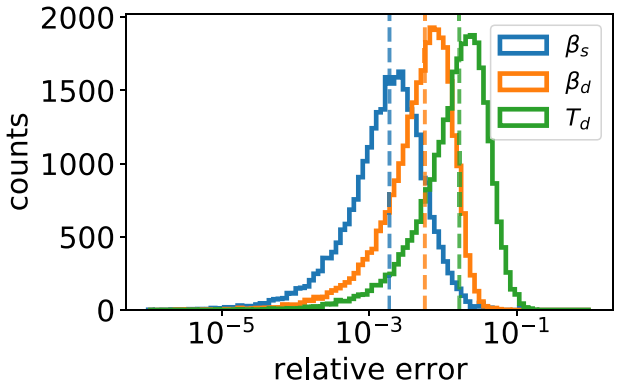

Figure 10. Histograms evaluated from the relative error maps of β_s, β_d, T_d (the right column of Fig. 8), respectively, in solid blue, orange, and green. Shown as vertical dashed lines the median values, namely 0.002, 0.006, 0.017, respectively, for β_s, β_d, T_d.

As the output maps of FGBuster are estimated on a partial sky (see Fig. 9), we estimate the power spectra in the observed regions with NaMaster (Alonso, Sanchez & Slosar 2019) to correct the power spectra for the $E - B$ leakage introduced by masking. We show in Fig. 11 the angular power spectra of CMB B-mode polarization anisotropies as estimated from the recovered map outputs of FGBuster.

We notice that the statistical residuals are not affected by the mis-modelling since the instrumental noise is the dominant contribution to the post-component separation residuals. On the other hand, the mis-modelling is clearly visible when we assess the systematics bias in the residual maps. In particular, we observe an increase of the systematic bias proportional to f, indicating how the partition is less and less representative with increasing values of f of the underlying Galactic foreground emission. However, even with the largest mis-modelling scenario $f = 1$, the residuals increase by only an order of magnitude. As expected, larger f_{sky} (including lower Galactic latitudes) yields larger residuals in the power spectra, this is the reason why we observe higher residuals for LiteBIRD with respect to the SO-SAT case).

Furthermore, to show the effective gain of performing parametric component with clustering, we consider two extreme cases of systematic residual estimates. We mainly focus on LiteBIRD channels and, for these specific tests no clustering methodology is employed. We simply skip the fitting procedure and consider two sets of maps of β_d, β_s, T_d with $f = 0$ and $f = 1$ (see equation 10), each set been provided as solution to the component separation problem. This test results in delivering a *range of allowance*, where we expect clustering to perform between the lower and upper limit. The lower limit [not shown in Fig. 11 (left) for graphical purposes] represents the best case scenario: the parameters of the model and the ones adopted for LiteBIRD observations exactly match (since $f = 0$). This test gives numerically zero residuals as the maps provided to the parametric fit are the exact solution of the problem (i.e. the same spectral parameter maps with which we have constructed the frequency channels). Consequently, the B-mode spectrum amplitude at every ℓ is numerically close to 0. The upper limit of the grey-shaded area in Fig. 11 (left) is instead estimated by providing maps of β_d, β_s, T_d with $f = 1$, hence fully mis-modelled with respect to the templates adopted to simulate the LiteBIRD frequency channels. This represents the worst case scenario where the mis-modelling propagates fully through the parametric fitting pipeline for each pixel of the map. This results in a residual mis-modelling bias of $\sim 5\,\mu\mathrm{K}^2_{\mathrm{CMB}}$ almost independent on ℓ. The fact that the power spectra obtained with clustering are well within this range, quantifies the overall benefit of performing the parametric fitting by partitioning the sky with respect to the pixel-by-pixel case, e.g. compare the case with clusters are obtained with the largest mis-modelling case ($f = 1$) with the upper limit of the shaded grey area in Fig. 11 (left).

Finally, we notice in Fig. 11 that all the B-mode systematic residual spectra tend to converge to a certain residual level at around $\ell \gtrsim 60$. Hereafter, we refer to it as a *partition noise* being essentially related to the combination of the typical angular scales of the cluster patches. This results as a 'noise' residual term in the power spectrum, as that prevents to estimate the variability of the spectral parameters at multipoles larger than this scale.

4.2 Clustering applied on H I clouds maps.

In the previous section, the component separation is performed on partitions derived from the same maps used in the parametric modelling, i.e. β_s, β_d, T_d. However, this relies on the assumption that spectral parameter maps, from which we derive the cluster patches, are not contaminated by other Galactic (or extra-galactic) foregrounds. Indeed, this is commonly the case for both the thermal dust emission, mainly affected by the cosmic infrared background (CIB) residuals and for synchrotron contaminated by free–free and AME.

Moreover, given the fact that the use of ancillary data sets for foreground modelling has been proposed by a number of works (e.g. HI4PI Collaboration et al. 2016; Clark 2018), we wish to explore whether such an ancillary data set can in principle yield accurate enough results for CMB parameter estimation purposes. To this end,

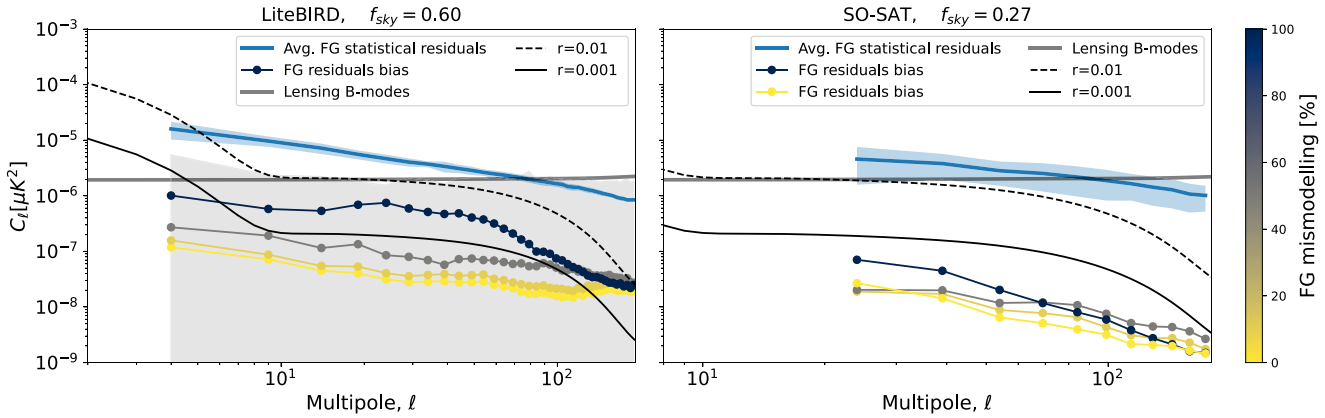

Figure 11. *B*-mode angular power spectra for the (left) LiteBIRD and (right) SO-SAT cases, estimated with NaMaster on to the recovered CMB maps from FGBuster performed on to cluster regions. (solid blue thick) Average of foreground statistical residuals obtained from 20 MC signal+noise simulations (shaded blue) 1σ standard deviation of MC simulations, (filled circles) foreground residual bias estimated with noiseless component separation for different choices of foreground mis-modelling given by the colourbar. The shaded grey area in the left-hand panel indicates two extreme cases: the upper limit is obtained by propagating through FGBuster the $f = 1$ mis-modelling to the spectral parameters, the lower limit spectra (not shown here) is nearly equal to zero and it is estimated by propagating in the component separation the $f = 0$ case. As a reference we show the primordial B modes for two different chosen values of tensor-to-scalar ratio, $r = 0.01, 0.001$, respectively, in dashed black and solid black. Lensing B modes are also shown in solid thick grey. The spectra in the left-hand and right-hand panels are binned, respectively, with $\Delta\ell = 5$ and $\Delta\ell = 15$.

we use a complementary data set that traces foreground dust: H I emission. In particular, we make use of the parametrization of line-of-sight complexity in the dust distribution presented by Panopoulou & Lenz (2020). We briefly explain the data set and our post-processing of it in the following.

Panopoulou & Lenz (2020) measured the number of clouds along the line of sight that appear as distinct peaks in H I spectra. They present a peak-finding algorithm that makes use of the HI4PI survey (HI4PI Collaboration et al. 2016) over the high Galactic latitude sky and outputs a measure of the number of clouds in each pixel. They introduced a robust measure of the number of clouds expected to contribute to the dust emission, which takes into account the relative contribution of the dust from each cloud along the line of sight, defined as $\mathcal{N}_c$:

$$\mathcal{N}_c = \sum_{i=1}^{N_{\rm clouds}} \frac{N^i_{\rm H\,I}}{N^{\rm max}_{\rm H\,I}}, \tag{11}$$

where $N^i_{\rm H\,I}$ refers to the column density of the ith cloud along the sightline and $N^{\rm max}_{\rm H\,textsci}$ is the highest value of column density for a cloud in the same direction. We therefore expect non-integer values for $\mathcal{N}_c$ especially when there is an imbalance in the column density for clouds along a sightline. The $\mathcal{N}_c$ map was estimated by segmenting the sky into large *superpixels* at `nside=128` (corresponding to ∼30 arcmin), with each superpixel encoding 64 pixels of the HI4PI map. Recently, Pelgrims et al. (2021) adopted the maps publicly released by Panopoulou & Lenz (2020) to estimate the line-of-sight frequency decorrelation of dust polarization in the *Planck* maps. For the purposes of our clustering analysis, we require maps at `nside=32` and `nside=64`. We thus repeat the analysis in Panopoulou & Lenz (2020) to produce maps at these lower `nsides`, instead of downgrading the publicly available maps, as recommended by the authors. The publicly available $\mathcal{N}_c$ maps did not include uncertainties. As our clustering relies on the use of uncertainties, we calculate the per-pixel uncertainty in the $\mathcal{N}_c$ maps that we use as described in Appendix D.

As discussed in Panopoulou & Lenz (2020, Section 5.3.1), the derived maps do not present any imposed spatial coherence at scales above the superpixel size, leading to the eventuality of discontinuities between neighbouring pixels. However, clustering methodologies presented in this work can implement the spatial coherence above a given pixel scale.

We employ the $\mathcal{N}_c$ and the associated uncertainties as in equation (D2) with maps at `nside=32` (shown in top panel of Fig. 12) and run the spectral clustering methodology outlined in Section 3. We then evaluate the total variance of the partition in a similar grid to the one shown in Fig. 6 and find the minimum variance to be at $\alpha = 0.10$ and $\delta = 0.20$, corresponding to $K = 1481$ clusters.

To evaluate how well the clustering morphologies track the input map, we estimate the median value of pixels of $\mathcal{N}_c$ belonging to the same clusters and produce a *binned* map, $\mathcal{N}_c^{bin}$. We can then estimate the relative error defined as

$$\delta\mathcal{N}_c = |\mathcal{N}_c - \mathcal{N}_c^{bin}|/\mathcal{N}_c. \tag{12}$$

The bottom row of Fig. 12 shows in the left-hand panel the partition estimated with spectral clustering and in the right one the relative errors. We observe that especially around the North and South Galactic Poles ($90^\circ < |b| < 60^\circ$) the clustering optimization tends to produce smaller clusters, essentially encoding few pixels per cluster. On the contrary, at intermediate Galactic latitudes, clusters contain a larger number of pixels. This might be related to the fact that the $\mathcal{N}_c$ maps are produced on too large pixelization (`nside=32`), resulting in an overall reduction of the uncertainty budget and a consecutive increase in SNR of the estimates. As already stated in Section 4.1, in the presence of very high SNR features, the optimal partition leads to *pixel-sized* clusters. We will devote a further investigation on this respect in a future work.

Notice that in the regions where the uncertainties are small, the error is below ∼ 10 per cent. Vice versa, in large uncertainties regions the relative error can be as much as 50 per cent although those regions are localized mainly in the North Galactic Hemisphere. However, we remark here that these larger errors should not be compared with the ones shown in Fig. 8 as larger errors on $\mathcal{N}_c$ do not necessarily translate on larger residuals on the recovered dust parameters.

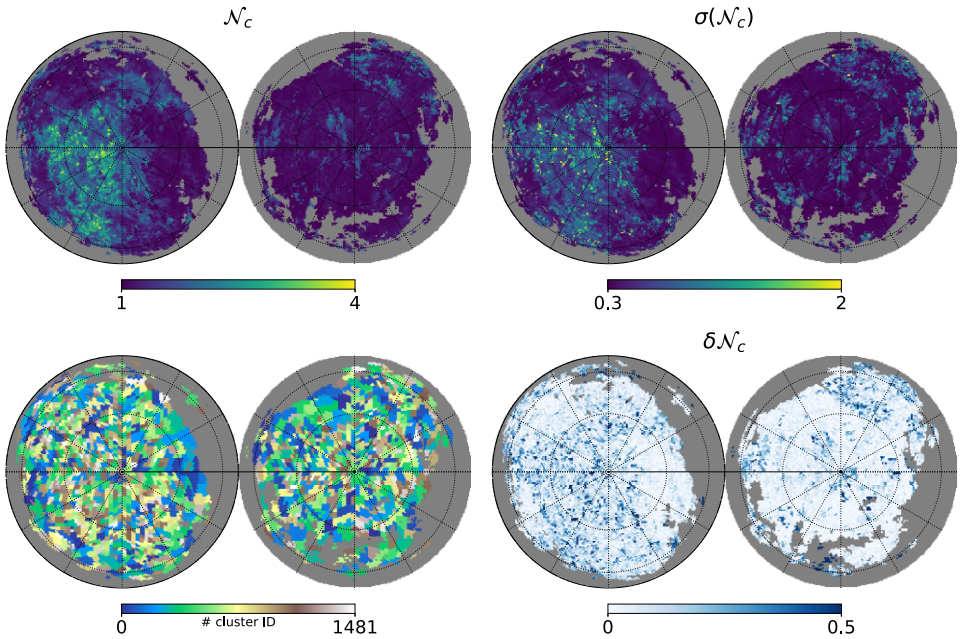

Figure 12. Top: Maps of $\mathcal{N}_c$ (left) and its uncertainties $\sigma(\mathcal{N}_c)$ (right) evaluated as in equation (D2). Bottom left: Partition of $\mathcal{N}_c$; each colour identifies a different clustering region. Bottom right: Relative errors $\delta\mathcal{N}_c$ estimated with (equation 12).

As already mentioned in Panopoulou & Lenz (2020), one of the interesting applications of the map of number of clouds per line of sight is to inform realistic models of polarized dust emission for parametric component separation methods as the one we adopt in Section 4.1. In this specific case, we utilize the patches derived from the $\mathcal{N}_c$ map (the bottom left-hand panel in Fig. 12) to perform the component separation estimation of dust and CMB (simulated at `nside=64`). In fact, since the $\mathcal{N}_c$ map traces the dust emission, we do not include the synchrotron as well as all the $\nu < 90$ GHz frequency channels in this analysis. This helps in better singling out the effects of residuals due to the dust component only.

Furthermore, to run NaMaster we regularize the HI4PI observation area ($f_{sky} = 0.55$ per cent) by apodizing it, yielding to ~ 0.35 per cent of the sky. For the run with SO-SAT we combine this mask with the SO-SAT one shown in Fig. 9 (bottom). The combined observation patches results in a further reduction, $f_{sky} \sim 22$ per cent.

In Fig. 13, we show the B-mode power spectra of the residuals in the recovered CMB map estimated with NaMaster. We observe that the statistical residuals are reduced by about one order of magnitude compared to the ones shown in Fig. 11. This is mostly due to the fact that we have excluded the low frequency channels that have lower sensitivity (see Table 1), yielding larger post-component separation noise when taken into account. Interestingly, the residual bias achieves levels comparable to $r \lesssim 10^{-3}$. It therefore seems that the parameters from the component separation performed on the dust-emission-agnostic $\mathcal{N}_c$-based sky partitioning are accurately representing the underlying signal.

One might wonder whether this dust-emission-agnostic clustering constitutes a real benefit in parameter estimation or it is only related to the fact that we are simply partitioning the northern and southern polar caps with small patches. In other words: *is there intrinsic value in the information contained in the number-of-clouds map?*

To answer this question, we disentangle the sky partition from the underlying physical information by flipping the cluster coordinates north to south. However, since the north and south Polar caps have different footprints, we firstly flip the map and then apply the regularized mask to the $\mathcal{N}_c$ map. Encoding the mask on a smaller fraction of sky ensures that each pixel in the flipped map falls within the $\mathcal{N}_c$ map footprint.

We then perform the parametric fit with the same combination of frequency channels as before but with the flipped cluster regions. The thick dashed lines shown in Fig. 13 indicate the power spectra of systematic residuals from the output of this component separation run and we notice a remarkable increase of the residuals especially at lower multipoles both for SO-SAT and LiteBIRD.

A visual inspection of the relative error and standard deviation maps estimated in both cases for T_d and β_d, reveals that both the error and the standard deviation (particularly in T_d) increase when we consider the case with flipped regions. This also indicates that the bias we observe in Fig. 13 is mostly due to wrong estimation of the T_d parameter.

This is not unexpected since $\mathcal{N}_c$ has been shown to be correlated with T_d (Panopoulou & Lenz 2020). The higher values of $\mathcal{N}_c$ coincide with regions where the emission from two physically distinct families of clouds overlap (low velocity and intermediate velocity clouds). One family of clouds resides at larger distances from the Galactic plane and is found to have higher dust temperatures, presumably due to dust shattering as it falls on to the mid-plane (Planck Collaboration XXIV 2011). Therefore, a mis-modelling of the cloud population (e.g. by flipping the north–south map) naturally leads to a wrong estimate of T_d. We thus can conclude that partitioning the sky observations with patches inferred from the $\mathcal{N}_c$ map can improve the performances of parametric fitting methodologies.

Finally, we observe also for this case the effects of a partition noise in the power spectra: with B-mode residuals in Fig. 13 (dashed lines and filled circles) converging at around $\ell \sim 60$. A partition with smaller patches, as the one shown in Fig. 7, results into shifting to higher multipoles the contribution due to this partition noise. Grumitt et al. (2020) proposed a hierarchical approach to overcome exactly this limitation and we plan to integrate it in a future work.

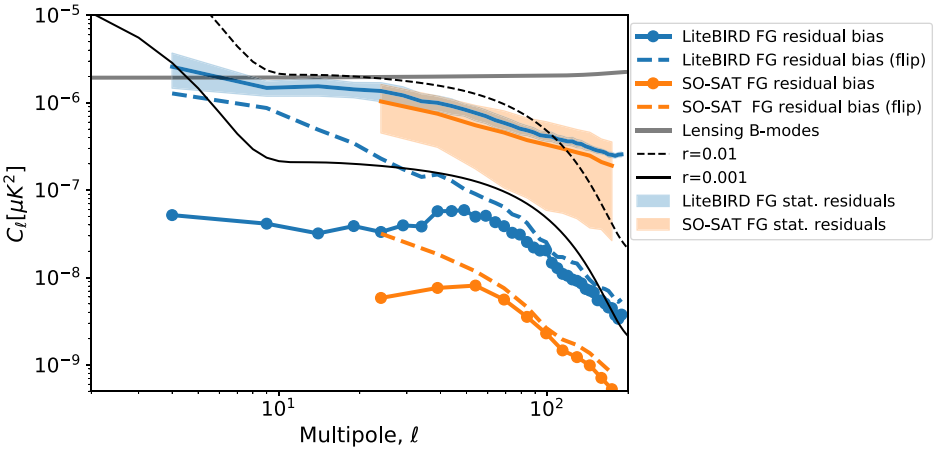

Figure 13. B-mode residuals from the recovered CMB maps output of FGBUSTER performed within multiple regions defined by clustering of $\mathcal{N}_c$ map for (blue) LiteBIRD and (orange) SO-SAT frequency channels, respectively, on $f_{sky} = 0.35$ and 0.22. (filled circles) Foreground residual bias performed on noiseless simulations, (shaded area) 1σ standard deviation of statistical residuals evaluated from 20 MC signal and noise simulations (the average of the 20 MC realization is shown as solid line). (dashed) Foreground residual bias estimated from the component separation run performed by flipping the map of $\mathcal{N}_c$ clusters north to south. These spectra show an excess at lower multipoles, demonstrating that the $\mathcal{N}_c$ map provides useful constraints for parameter estimation.

4.3 Estimates on *r*

We recall the reader that to specifically assess the bias introduced by foregrounds, we do not include the CMB emission in the simulated frequency maps. However, we instead fit for CMB in the parametric component separation. Given the B-mode power spectra shown in Figs 11 and 13, we estimate the value of residual bias in terms of r by evaluating the likelihood $\mathcal{L}(r)$ in the binned multipole domain, ℓ_b, given the fact that the power spectra have been corrected by the mode coupling with NAMASTER.

The likelihood function reads as

$$\ln \mathcal{L}(r) = \sum_{\ell_b=\ell_{\rm min}}^{\ell_{\rm max}} \ln\left(-f_{sky}\Delta\ell\frac{2\ell_b+1}{2}\left[\frac{\hat{C}_{\ell_b}}{C_{\ell_b}} + \ln C_{\ell_b}\right]\right), \quad (13)$$

where $\ell_{\rm max} = 200$, $\ell_{\rm min} = 2\,(30)$, and $\Delta\ell = 5(15)$ for LiteBIRD (SO-SAT) and $\hat{C}_{\ell_b}$ (C_{ℓ_b}) is the measured (modelled) B-mode power spectrum. As we are interested in estimating the bias of systematic residuals we assume the measured B-mode spectrum to encode all the contributions but the primordial one ($r = 0$), i.e.

$$\hat{C}_{\ell_b} = C^{\rm lens}_{\ell_b} + \left\langle C^{\rm fg,tot}_{\ell_b}\right\rangle, \quad (14)$$

where $\langle C^{\rm fg,tot}_{\ell_b}\rangle$ is the expected statistical noise power spectrum residuals after component separation obtained by averaging the spectra of 20 MC signal+noise simulations (Errard & Stompor 2019) and $C^{\rm lens}_{\ell_b}$ is the lensing B-mode power spectrum. The modelled power spectrum is given as

$$C_{\ell_b} = rC^{\rm tens}_{\ell_b} + C^{\rm lens}_{\ell_b} + \left\langle C^{\rm fg,tot}_{\ell_b}\right\rangle - C^{\rm fg,sys}_{\ell_b}, \quad (15)$$

where $C^{\rm tens}_{\ell_b}$ is the tensor mode with $r = 1$ and $C^{\rm fg,sys}_{\ell_b}$ is the spectrum encoding the systematics bias due to mis-modelling the foreground emission.

Thus, the bias on r, Δr is defined as the value that maximizes the likelihood function:

$$\left.\frac{{\rm d}\mathcal{L}(r)}{{\rm d}r}\right|_{r=\Delta r} = 0, \quad (16)$$

Table 2. Estimated values for the bias and the uncertainty on r obtained by evaluating the likelihood in equations (13), (16), and (17).

Clustering on β_s, β_d, T_d FG Mis-modelling f	LiteBIRD	SO-SAT
0	$(0.22 \pm 0.66) \times 10^{-3}$	$(0.08 \pm 1.21) \times 10^{-3}$
0.1	$(0.28 \pm 0.66) \times 10^{-3}$	$(0.10 \pm 1.21) \times 10^{-3}$
0.5	$(0.59 \pm 0.67) \times 10^{-3}$	$(0.14 \pm 1.21) \times 10^{-3}$
1.0	$(2.18 \pm 0.68) \times 10^{-3}$	$(0.20 \pm 1.19) \times 10^{-3}$
Clustering on $\mathcal{N}_c$	$(0.09 \pm 0.26) \times 10^{-3}$	$(0.04 \pm 0.16) \times 10^{-3}$

and the error on r, δr is defined as the value covering the 68 per cent area of the total likelihood function, i.e.

$$\frac{\int_0^{\delta r}\mathcal{L}(r){\rm d}r}{\int_0^{\infty}\mathcal{L}(r){\rm d}r} = 0.68. \quad (17)$$

The values reported in Table 2 are extensively discussed in the following section.

5 DISCUSSION

We devote this section to discuss in more detail the results presented in Section 4. We estimate how the bias from systematic residuals for different choices of the mis-modelling parameter f affects the detection of primordial CMB B modes and compare with the requirements for SO-SAT and LiteBIRD.

As reported in Table 2 and shown in Fig. 14, the effect of foreground mis-modelling is visible in terms of systematic bias on r estimate but slightly affects the uncertainties (which are instead driven by the sensitivity of the experiment, $f_{\rm sky}$, frequency coverage, etc). Although the bias for LiteBIRD is always $<10^{-3}$ for most cases, we observe an increase proportional to the mis-modelling. Without further optimizing the Galactic mask for polarization the values in the top row of Table 2 show that we can meet the latest LiteBIRD requirements ($\delta r < 10^{-3}$, Sugai et al. 2020) with a mis-modelling $f \leq 0.5$.

We notice that the bias on r obtained in this work for SO-SAT is slightly lower than the reported values in Simons Observatory

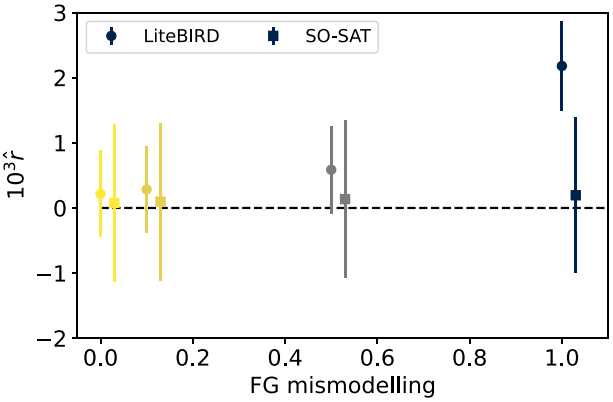

Figure 14. r estimates and the accompanying uncertainties evaluating the likelihood in equations (13), (16), and (17). We adopt the same colour scheme as in Fig. 11 to indicate different amounts of mis-modelling.

Collaboration et al. (2019, table 4), indicating a net improvement thanks to this technique. The larger uncertainties and the lower systematic bias observed in SO-SAT, with respect to LiteBIRD ones, are mostly due to the smaller coverage both in frequency and in the angular scales observed. Moreover, the r estimates for SO-SAT are assessed by evaluating the likelihood in a limited range of ℓ ($30 < \ell < 200$). Vice versa, LiteBIRD benefits of the full-sky angular range allowing also the large angular scales to be included in the likelihood ($2 < \ell < 200$).

Alternatively to foreground parameter maps, we also apply clustering to partition the map of number of dust clouds along the line of sight, $\mathcal{N}_c$, presented in Panopoulou & Lenz (2020). We perform the same parametric fitting as before on the patch derived from $\mathcal{N}_c$, and find similar level of residuals post-component separation for both SO-SAT and LiteBIRD cases when compared with ones obtained from clusters derived with foreground parameter maps. Not only do we find that the partition obtained is representative of the dust emission, but also that the patches inferred from $\mathcal{N}_c$ yield lower residuals in the estimates of the thermal dust temperature parameter, T_d.

We further estimate the leakage in terms of r also for this case of application by means of the likelihood defined in equation (13) and we found $\hat{r} = (0.09 \pm 0.26) \times 10^{-3}$ and $(0.04 \pm 0.16) \times 10^{-3}$, respectively, for LiteBIRD and SO-SAT. The reason we obtain lower biases with respect to the one reported in Table 2 for the $f = 0$ mis-modelling case is mainly due to the fact that low-frequency channels are not accounted for in the component separation and we assume the synchrotron to be negligible at $\nu > 90$ GHz. Moreover, we get a lower noise bias for SO-SAT with respect to LiteBIRD mainly because to the fact that we do not account for the $1/f$ noise in our simulations for both instruments. However, the low bias on r is a stricking indication that the $\mathcal{N}_c$ map can be a very promising tracer to probe the thermal dust emission in the range of 100−300 GHz. Although Pelgrims et al. (2021) showed that the $\mathcal{N}_c$ maps can be used as a tracer for frequency decorrelation in the dust polarization map, the results shown in this paper offer a further indication of their potential.

As already mentioned, this is not the first time that clustering methodologies are employed to segment the sky into patches where to perform foreground cleaning [see Grumitt et al. (2020) where they used the mean-shift algorithm]. However, by comparing the maps of clusters in Fig. 7 with those in Fig. 2 of (Grumitt et al. 2020), we would like to address several reasons why the two methods yield very different partitions.

One immediate consideration is the fact that mean-shift and spectral clustering are based on different algorithms that naturally might yield different results even when applied on samples with fewer features than the ones in this context.

The clusters obtained in Grumitt et al. (2020) are larger than the ones shown in Fig. 7, although we also notice a trend in estimating larger clusters particularly for the β_s clusters. This makes the method better suited for low-resolution and low-frequency channels where synchrotron emission dominates,[12] but it was essentially unable to track the finer structure of the dust parameters.

One of the major differences is that we derive the clusters *separately* for each spectral parameter, whereas in Grumitt et al. (2020) the dust and synchrotron spectral indices β_s and β_d are combined together to derive a single cluster map used to perform the parametric fit (see fig. 2 of Grumitt et al. 2020).

Furthermore, the features in Grumitt et al. (2020) included the two spectral indices together with the Cartesian coordinates on the sphere x, y, z. On the contrary, our analysis does not rely on the coordinates as features since the adjacency is defined from the distance between pixels (Section 3.1). We further account also for T_d as an extra-parameter for the thermal dust emission, leading to a reduction in the foreground bias especially at high Galactic latitudes.

A novel feature of the technique presented in this work is to perform the clustering by accounting for the statistical significance of the measurements reported in the spectral parameter maps and their uncertainties. We weigh the pixel adjacency in such a way that the significance plays a non-negligible role in defining the pixel similarities especially in regions of high SNR. Moreover, we drastically reduce the dimensionality of our problem by performing the clustering on the eigenspace spanned by the eigenvectors related to the smallest 256 eigenvalues. This avoids high-dimensionality problems occurring frequently in clustering algorithms with large number of features [e.g. the number of pixels in a `nside=64` HEALPix map see section 4 of Grumitt et al. (2020)].

Finally, in Appendix B, we report also the clustering analysis performed on the GNILC β_d and T_d maps. The morphologies are slightly different with respect to the ones observed in the clusters in Fig. 7 as the spectral parameter features and the uncertainties in the two data sets are different.

6 CONCLUSIONS

In this work, we aim at identifying regions in the sky within which the Galactic foreground emission can be assumed homogeneous. We define a partitioning of the full celestial sphere by means of a novel technique based on a spectral clustering algorithm embedded in the S^2 manifold. In contrast to previous applications of clustering for CMB studies, this method works directly on the celestial sphere instead of mapping points on to a Cartesian grid. It also takes into account the uncertainties in the clustering procedure. To our knowledge, this is the first time where image segmentation has been performed on features defined on the S^2 manifold.

Another key advancement of our algorithm is the use of an objective criterion for selecting the optimal partition parameters (α and δ). Many *off-the-shelf* algorithms rely on subjective measures, such as visual inspection, to select a 'reasonable' segmentation of the

[12] An interesting application in Grumitt et al. (2020) was indeed the combination of the C-BASS data at 5 GHz with the LiteBIRD frequency channels.

domain (e.g. the maximum pixel size parameter in mean-shift). The algorithm presented here borrows established metrics for partition measure evaluation in clustering problems to produce an optimal partition. We rely on the comparison of within- and between-cluster variances for this, while adding new modifications to incorporate the presence of uncertainties in the data.

We apply this clustering approach to evaluate Galactic contamination in CMB polarization studies. We firstly characterize the emission of thermal dust and synchrotron available from the latest observations of the spectral parameter maps (together with their uncertainties) to divide the celestial sphere into regions with spectral clustering. The choice of using these maps of spectral parameters is mainly due to the fact that those are the template employed in PySM to simulate the foreground spatial variability. However, since they are released at low-resolution, we applied the clustering to maps pixelized on a coarser Healpix grid (`nside=32`). With the future CMB surveys, e.g. SO, CMB-S4, we will be able to estimate high-resolution templates and with an unprecedented sensitivity both in intensity and polarization. Moreover, several methodologies in the literature (Khatri 2015; Grumitt et al. 2020) aimed at estimating spectral parameters at high-resolution. We devote a future work to employ similar methodologies in the context of parametric fitting. We then perform a parametric component separation for each region simulating the nominal frequency channels from two forthcoming CMB experiments observing from the ground and from space. The optimal clustering partition yields low residuals in the CMB polarization maps after component separation, even in the case where the Galactic emission is mis-modelled up to some extent.

We further apply clustering to partition the map of number of HI clouds along the line of sight. Although this is one of the first applications using this kind of tracer, we find that clusters derived from the $\mathcal{N}_c$ map yield a residual bias as low as $r \sim 10^{-3}$ and that the dust spectral parameters are estimated on these regions up to < 10 per cent relative errors. This is a remarkable result since the constraints on T_d from far-infrared data have been so far complicated by the degeneracy with the spectral index β_d. On the other hand, the use of ancillary data combined with the latest data-science techniques shows promising results in order to improve the modelling of Galactic foregrounds.

In conclusion, this technique could be applied on a wide range of contexts involving spherical images, e.g. Earth images, identification of Solar features, wide astronomical surveys and it can be easily extended to higher dimensional data sets, like 3D surveys or cosmological simulations.

ACKNOWLEDGEMENTS

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. Giuseppe Puglisi acknowledges support by the COSMOS & LiteBIRD networks of the Italian Space Agency. The authors thank Clement Leloup, Carlo Baccigalupi, for having read the paper thoroughly.

Giuseppe Puglisi would like to thank: Jonathan Aumont, Mathieu Remazailles, Jens Chluba, Aditya Rotti, Susanna Azzoni, Leo Vacher, Hans Christian Eriksen, Nicoletta Krachmalnicoff for useful comments and discussions.

Georgia Panopoulou acknowledges support for this work by NASA through the NASA Hubble Fellowship grant #*HST*-HF2-51444.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

DATA AVAILABILITY

The whole clustering analysis presented here has been collected into a `python` package, ForeGround Clusters (FGCLUSTERS https://github.com/giuspugl/fgcluster[13]).

The inputs used throughout this paper together with the outputs obtained with the Clustering technique have been made publicly available online[14] and in the Harvard Dataverse:

(i) $\mathcal{N}_c$ map : https://doi.org/10.7910/DVN/XAMJ4X;
(ii) PySM spectral parameters: https://doi.org/10.7910/DVN/WJUEFA;
(iii) GNILC dust parameters: https://doi.org/10.7910/DVN/02SCHB.

REFERENCES

Alonso D., Dunkley J., Thorne B., Næss S., 2017, Phys. Rev. D, 95, 043504
Alonso D., Sanchez J., Slosar A., 2019, MNRAS, 484, 4127–4151
Azzoni S., Abitbol M., Alonso D., Gough A., Katayama N., Matsumura T., 2021, Journal of Cosmology and Astroparticle Physics, 2021, 047
Basak S., Delabrouille J., 2011, MNRAS, 419, 1163
Belkin M., Niyogi P., 2003, Neural Comput., 15, 1373
Bennett C. L. et al., 1992, ApJ, 396, L7
Berry T., Giannakis D., 2020, Commun. Pure Appl. Math., 73, 689
Berry T., Sauer T., 2016, Appl. Comp. Harmonic Anal., 40, 439
Bianchini F. et al., 2020, ApJ, 888, 119
Bobin J., Moudden Y., Starck J. L., Fadili J., Aghanim N., 2008, Stat. Meth., 5, 307
Boyd Z., Bae E., Tai X., Bertozzi A. L., 2018, SIAM J. Appl. Math., 78, 2439
Chluba J., Hill J. C., Abitbol M. H., 2017, MNRAS, 472, 1195
Choi S. K. et al., 2020, JACP, 12, 045
Clark S. E., 2018, ApJ, 857, L10
Coifman R., Lafon S., 2006, Appl. Comp. Harmonic Anal., 21, 5
Delabrouille J., Cardoso J. F., Patanchon G., 2003, MNRAS, 346, 1089
Dunkley J. et al., 2009, AIP Conf. Proc., 1141, 222
Eriksen H. K. et al., 2006, ApJ, 641, 665
Eriksen H. K., Jewell J. B., Dickinson C., Banday A. J., Górski K. M., Lawrence C. R., 2008, ApJ, 676, 10
Errard J., Stompor R., 2019, Phys. Rev. D, 99
Errard J., Stivoli F., Stompor R., 2011, Phys. Rev. D, 84, 063005
Fredholm I., 1903, Acta Math., 27, 365
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Grumitt R. D. P., Jew L. R. P., Dickinson C., 2020, MNRAS, 496, 4383
Guth A. H., 1981, Phys. Rev. D, 23, 347
H. Hu J. S., Bertozzi A. L., 2015, Proceedings of 10th International Conference, EMMCVPR, Hong Kong, Energy Minimization Methods in Computer Vision and Pattern Recognition. Hong Kong, China
Hansen F. K., Banday A. J., Eriksen H. K., Górski K. M., Lilje P. B., 2006, ApJ, 648, 784
Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, A&AS, 47, 1
HI4PI Collaboration B. et al., 2016, A&A, 594, A116
Hinshaw G. et al., 2009, ApJS, 180, 225
Hu W., White M. J., 1997, New Astron., 2, 323
Irfan M. O., Bobin J., Miville-Deschênes M.-A., Grenier I., 2019, A&A, 623, A21
Khatri R., 2015, MNRAS, 451, 3321
Khatri R., 2019, J. Cosmol. Astropart. Phys., 2019, 039

[13] https://github.com/giuspugl/fgcluster
[14] https://portal.nersc.gov/project/sobs/users/giuspugl/Clustering

Krzanowski W. J., Lai Y. T., 1971, in Marriot, ed., A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares. Vol. 27, p. 501
Leach S. M. et al., 2008, A&A, 491, 597
M. Reuter F.-E. W., Peinecke N., 2005, SPM '05: Proceedings of the 2005 ACM Symposium on Solid and Physical Modeling. Cambridge, Massachusetts, p. 101
Maino D., Donzelli S., Banday A. J., Stivoli F., Baccigalupi C., 2007, MNRAS, 374, 1207
Mangilli A., Aumont J., Rotti A., Boulanger F., Chluba J., Ghosh T., Montier L., 2021, A&A, 647, A52
Meng Z., Merkurjev E., Koniges A., Bertozzi A. L., 2017, Hyperspectral Image Classification Using Graph Clustering Methods, Image Processing Online. Vol. 7, p. 218
Mihaylov G., Spallanzani G., 2016, International Journal of Prognostics and Health Management, Special Issue Big Data and Analytics, Emergent Behaviour in a System of Industrial Plants Detected via Manifold Learning. Vol. 7
Miville-Deschênes M.-A., Ysard N., Lavabre A., Ponthieu N., Macías-Pérez J. F., Aumont J., Bernard J. P., 2008, A&A, 490, 1093
Panopoulou G. V., Lenz D., 2020, ApJ, 902, 120
Pelgrims V., Clark S. E., Hensley B. S., Panopoulou G. V., Pavlidou V., Tassis K., Eriksen H. K., Wehus I. K., 2021, A&A, 647, A16
Planck Collaboration XXIV, 2011, A&A, 536, A24
Planck Collaboration XII, 2014, A&A, 571, A12
Planck Collaboration IX, 2016a, A&A, 594, A9
Planck Collaboration X, 2016b, A&A, 594, A10
Planck Collaboration X, 2016c, A&A, 594, A10
Planck Collaboration XIII, 2016d, A&A, 594, A13
Planck Collaboration XLVIII, 2016e, A&A, 596, A109
Planck Collaboration IV, 2020a, A&A, 641, A4
Planck Collaboration VI, 2020b, A&A, 641, A6
Puglisi G. et al., 2018, A&A, 618, A62
Puglisi G., Fabbian G., Baccigalupi C., 2017, MNRAS, 469, 2982
Remazeilles M. et al., 2018, J. Cosmol. Astropart. Phys., 2018, 023
Rustamov R. M., 2007, Proc. Eurograph. Symp. Geometry, 15, 225
Seljak U., Zaldarriaga M., 1997, Phys. Rev. Lett., 78, 2054
Simons Observatory Collaboration et al., 2019, J. Cosmol. Astropart. Phys., 2019, 056
Singer A., Wu H.-T., 1997, Inf. Infer. J. IMA, 6, 58
Starobinsky A. A., 1982, Phys. Lett. B, 117, 175
Stolyarov V., Hobson M. P., Lasenby A. N., Barreiro R. B., 2005, MNRAS, 357, 145
Stompor R., Leach S., Stivoli F., Baccigalupi C., 2008, MNRAS, 392, 216
Stompor R., Errard J., Poletti D., 2016a, Phys. Rev. D, 94, 083526
Stompor R., Errard J., Poletti D., 2016b, Phys. Rev. D, 94, 083526
Sugai H. et al., 2020, J. Low Temp. Phys., 199, 1107
Szydlarski M., Grigori L. Stompor R., 2014, A&A, 572, A39
Tassis K., Pavlidou V., 2015, MNRAS, 451, L90
The Polarbear Collaboration et al., 2017, ApJ, 848,121
The Polarbear Collaboration et al., 2020, ApJ, 897, 55
Thorne B. et al., 2019, preprint (arXiv:1905.08888)
Thorne B., Dunkley J., Alonso D., Næss S., 2017, MNRAS, 469, 2821
Tristram M. et al., 2021, A&A, 647, A128
Von Luxburg U., 2007, Technical Report 4, A Tutorial on Spectral Clustering, www.springer.com
Wagner-Carena S., Hopkins M., Rivero A. D., Dvorkin C., 2019, MNRAS, 494, 1507
Zelnik-Manor L., Perona P., 2004, Technical Report, Self-Tuning Spectral Clustering, http://www.vision.caltech.edu/lihi/Demos/SelfTuningClustering.html
Zhang C., Zhu G., Chen M., Chen H., Wu C., 2018, Technical Report, Image Segmentation Based on Multiscale Fast Spectral Clustering, preprint (arXiv:1812.04816v1)
Zhao C., Song J. S., 2018, Front. Appl. Math. Stat., 4, 1
Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, J. Open Sour. Softw., 4, 1298

APPENDIX A: A GEOMETRIC INTUITION ON SPECTRAL IMAGE SEGMENTATION

The Laplace–Beltrami operator Δ_g is the generalization of the Laplace operator Δ (divergence of the gradient) on a Riemannian manifold with metric g. In Local coordinates, it can be written as

$$\Delta_g f = \frac{1}{\sqrt{|g|}}\partial_i\left(\sqrt{|g|}g^{ij}\partial_j f\right).$$

The Laplace–Beltrami operator is a second-order linear elliptic differential operator. It is the Hodge (connection) Laplacian acting on functions, it is by construction self-adjoint and positive-semidefinite.

The spectral properties of Δ_g on a compact differentiable manifold M are well known. The spectrum λ_i is discrete, $\lambda_1 = 0$ is always an eigenvalue, all eigenvalues $0 \leq \lambda_1 ... \leq \lambda_j \leq$ are positive. The eigenfunctions φ_i of the Laplace–Beltrami operator on M form a complete orthonormal basis of the space $L^2(M)$, i.e. any function in $L^2(M)$ can be written as a convergent series in $L^2(M)$ with real coefficients. This is the celebrated Sturm–Liouville decomposition of L^2 functions on smooth compact manifolds. The discussion of the eigenvalue problem on manifolds with boundary (with Dirichlet or Newmann boundary conditions) is often irrelevant in the discrete applications.

The Dirichlet energy functional of a function on M is defined as

$$E[f] = \frac{1}{2}\int_M ||\nabla f(x)||^2 \mathrm{d}\omega_g,$$

where ω_g denotes the volume form over a Riemannian manifold. The Dirichlet energy measures the 'variability' of f on M.

The Dirichlet functional evaluated on the eigenfunctions φ_i of Δ_g is monotonously increasing as the corresponding eigenvalues increase. The Stokes theorem applied on a function f and a vector field X implies that $-div$ and ∇ are formally adjoint operators (see Belkin & Niyogi 2003); that is,

$$\int_{\mathcal{M}}\langle X, \nabla f\rangle = -\int_{\mathcal{M}}(\nabla \cdot X) f.$$

So that

$$\int_{\mathcal{M}} \parallel \nabla f(x) \parallel^2 \mathrm{d}\omega_g = \int_{\mathcal{M}}\langle \nabla f, \nabla f\rangle \mathrm{d}\omega_g = -\int_{\mathcal{M}}(\nabla\cdot\nabla f) f \mathrm{d}\omega_g = -\int_{\mathcal{M}} f(\Delta_g f)\mathrm{d}\omega_g.$$

The former equation implies

$$E[\varphi_i] = \frac{1}{2}\int_M ||\nabla\varphi_i(x)||^2 \mathrm{d}\omega_g = \frac{1}{2}\int_M \langle\nabla\varphi_i, \nabla\varphi_i\rangle \mathrm{d}\omega_g = \frac{1}{2}\int_M \varphi_i(\Delta_g\varphi_i)\mathrm{d}\omega_g = \frac{1}{2}\lambda_i$$

For $f = \sum_i \alpha_i\varphi_i$ this expression linearly extends to

$$E[f] = \frac{1}{2}\sum_i \alpha_i^2\lambda_i.$$

The Laplace–Beltrami operator on $L^2(M)$ can be alternatively defined though the quadratic functional $E[f]$:

$$E[f] = \langle\nabla f, \nabla f\rangle_{L^2(M)} = \langle\Delta_g f, f\rangle_{L^2(M)}. \quad \text{(A1)}$$

Interestingly, $\nabla E[f] = \Delta_g f$ i.e. the Laplace–Beltrami operator is the variational derivative of Dirichlet energy in $L^2(M)$.

PROPOSITION 3. *The function that minimizes the Dirichlet energy and is orthogonal in* $\mathrm{L}^2(\mathrm{M})$ *to the space spanned by* $\varphi_0,..., \varphi_\mathrm{i}$ *is* $\varphi_{\mathrm{i}+1}$.

A1 Fredholm theory and discrete differential operators

Digitally produced images resemble smooth functions sampled discretely over regular domains (a grid of pixels), the above construction needs to be replicated in a discrete form. For some general ideas about the discretization of differential operators acting on sections of tensor bundles, we refer the reader to Mihaylov & Spallanzani (2016). Manifold learning techniques that include consistent discretizations of key elements of Riemannian geometry (vector fields, connections) have been described in Singer & Wu (1997). Recently this approach has been further developed in Berry & Giannakis (2020) including differential forms, Laplace-De Rham operators, etc.

Let us consider the following differential equation:

$$\mathcal{D}f = g, \tag{A2}$$

where $\mathcal{D} : L^2(M) \longrightarrow L^2(M)$ is a linear differential operator. The Green's function of a differential operator is a special integral kernel related to the Dirac's delta:

$$\mathcal{D}G(x, y) = \delta(x - y).$$

As an application of Fredholm's theory, the solution of equation (A2) can be written in the following equivalent integral form:

$$f(x) = \int_M G(x, y)g(y)\mathrm{d}\omega.$$

For operators with discrete spectrum and a complete orthonormal basis of eigenfunctions $\{\varphi_i\}$ the Green's function can be expressed:

$$G(x, y) = \sum_i \frac{1}{\lambda_i}\varphi_i(x)\varphi_i(y). \tag{A3}$$

If we consider a sampling of n points $x_i \in M$ and choose a conventional order for the sampled points, a finite sampling of a function f can be represented by a real n-dimensional vector. Denote by $G(x, y)$ the Green's function of a linear differential operator $\mathcal{D}$ on M. In line with Mihaylov & Spallanzani (2016), we call a *coarse discretization* of $\mathcal{D}$, the linear map $D : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined by

$$DG(x_i, x_j) = \delta_{ij}, \tag{A4}$$

where δ_{ij} denotes Kronecker's symbol. With this definition, the continuous eigenfuction problem $\mathcal{D}f = \lambda f$ is discretized into

$$\sum_i G(x_i, x_j)f(x_j) = \mu f(x_i), \tag{A5}$$

The Green's function measures the way in which two points are geometrically related. In the above problem, the contribution of the sampling $f(x_j)$ to the computation of $f(x_i)$ can be further localized by introducing an adjacency matrix W_{ij} that provides the sampling $x_i \in M$ with a graph structure. The elements of W_{ij} are equal to 1 or 0 based on different criteria depending on the problem, a distance-based cut-off, a fixed number of neighbouring points (pixels), etc. Equation (A5) becomes

$$Df(x_i) = \sum_{jk} G(x_i, x_j)W_{jk}f(x_k) = \mu f(x_i). \tag{A6}$$

Self-adjoint operators give rise to symmetric discretizations and the discrete version of equation (A3) is

$$G(x_m, x_n) = \sum_i \frac{1}{\lambda_i}\varphi_i(x_m)\varphi_i(x_n).$$

This equation and equation (A4) mean that D is the generalized inverse of G (or GW). By construction, if we increase the density of the sampling as $n \longrightarrow \infty$, the coarse discretization converges to the differential operator.

The coarse discretization of a known differential operator on a given manifold is a conceptually straightforward process and it differs from the typical manifold learning problem of approximating a linear differential operator $\mathcal{D}$ on a point cloud sampled from a priori unknown sub-manifold of $\mathbb{R}^n$. We call such a linear map the sample discretization of $\mathcal{D}$.

The elements used for building a sample discretization are the same as in equation (A6) i.e. an affinity graph and a kernel function $K(x_i, x_j)$ that satisfy certain symmetry and regularity assumptions (see Singer & Wu 1997; Berry & Sauer 2016). In the manifold learning set-up, the kernel $K(x_i, x_j)$ reproduces the expression of the (unknown) Green's function of the manifold in terms of the coordinates of the points x_i in the space of measured variables. A first approximation attempt regarding the expression of the kernel can be found between the Green's function of the same operator on $\mathbb{R}^n$ computed with the embedding coordinates x_i.

This estimate must be corrected by bias terms that include for example the curvature on M, the sample variance etc. Intuitively, the bias of the geodesic distance on M with respect to the Euclidean distance in the ambient space is measured by the curvature.

Given a sample discretization of a differential operator, two types of convergence problems arise in the limit of infinite sampling:

(i) Point convergence $D\varphi(x_i) \longrightarrow \mathcal{D}\varphi(x)$
(ii) Spectral – eigenvectors of D converge to eigenfunctions $\mathcal{D}$.

Spectral convergence is a stronger condition. Proving the convergence of a discretized operator is a non-trivial analytical problem that includes the curvature of the manifold, the density of sampling, normalization choices, etc. Remarkable results in this direction have been achieved in Belkin & Niyogi (2003) and Singer & Wu (1997).

A2 Discrete Laplacian

For a point cloud in $\mathbb{R}^n$, we define a graph with adjacency matrix A_{ij} with weights $K(x_i, x_j)$ and compute the (weighted) diagonal degree matrix $D_i = \sum_j A_{ij}$. There are several definitions of Laplace operators on graphs that use these elements. Typically, the adopted expressions for $K(x_i, x_j)$ are inspired by the Laplacian Green's function in the ambient space of the point cloud. For example the Green's functions of the Laplacian in $\mathbb{R}^2$ and $\mathbb{R}^3$ are well known:

$$G_{\mathbb{R}^2}(x, y) = \frac{1}{2\pi}\ln(|x - y|), \qquad G_{\mathbb{R}^3}(x, y) = \frac{1}{4\pi|x - y|}.$$

Berry & Sauer (2016) studied more general kernel constructions that take into account non-uniform sampling densities.

The most popular construction is the **graph Laplacian**

$$L := D - A. \tag{A7}$$

This definition arises from the Newton's cooling law on a graph. The heat transferred between connected nodes i and j is proportional to the difference $f(x_i) - f(x_j)$ and a heat capacity coefficient k, i.e.

$$\begin{aligned}\frac{\mathrm{d}f(x_i)}{\mathrm{d}t} &= -k\sum_j A_{ij}(f(x_i) - f(x_j)) \\ &= -k\left(f(x_i)\sum_j A_{ij} - \sum_j A_{ij}f(x_j)\right).\end{aligned}$$

This is further motivation for the degree matrix D in equation (A7).

The **random walk Laplacian** is defined as

$$L_{rw} := D^{-1}L = I - D^{-1}A \tag{A8}$$

and is related to Markov processes on graphs. The matrix L_{rw} is not symmetric, but row-stochastic (satisfies the Markov property). This definition has been further developed in the theory of diffusion maps (see Coifman & Lafon 2006; Singer & Wu 1997).

The **symmetric random walk Laplacian** is

$$L_{\rm sym} := D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}. \tag{A9}$$

The relations between these definitions and the transformations that map their spectral structures, respectively, are well known.

Remarkably, $L_{\rm sym}$ matrix approximates the Laplace–Beltrami operator both in pointwise and spectral sense (see Singer & Wu 1997; Belkin & Niyogi 2003). More precisely for $n \longrightarrow \infty$,

$$Lf(x_i) \approx \Delta f(x) + O\left(\frac{||\nabla f(x)||}{\sqrt{n}\epsilon^{1/2+dim(M)/4}}\right),$$

where ϵ is a parameter (see Berry & Sauer 2016). Moreover, in the case of uniform sampling, Belkin & Niyogi (2003) have shown that the eigenvectors of the graph Laplacian converge to the eigenfunctions of the Laplace–Beltrami operator on the manifold.

Constructions of discrete Laplacian operators from non-uniform distributions of manifold have been developed. Sample density can be absorbed in the definition of the kernel itself or a 'right normalization' is introduced Berry & Sauer (2016). These constructions are less relevant for the purposes of image segmentation, where images are usually sampled on regular girds of pixels.

The geometric relation between kernels and Riemannian metrics has been investigated in Berry & Sauer (2016) for uniform and non-uniform sampling. The expression of the Laplace–Beltrami operator changes if the Riemannian metric on M varies in a family of metrics, which includes the one induced by the embedding $M \hookrightarrow \mathbb{R}^n$. These changes are captured by modifications in the form of the Green's function of the operator and subsequently in its discretization. Every symmetric local kernel with exponential decay corresponds to a Laplacian operator in a Riemannian geometry and vice versa.

Once a discrete Laplacian is provided, the discrete Dirichlet energy functional of equation (A1) applied on a sampling $f(x_i)$ becomes

$$E[f(x_i)] = \frac{1}{2}\langle f(x_i), Lf(x_i)\rangle.$$

The construction of low-energy embedding of a graph in $\mathbb{R}^m$ with m lower or equal to the number of sampled points is directly applied in the discrete case and is used in spectral image segmentation.

A3 Spectral segmentation by means of the heat propagator

The geometric affinity/similarity between points on a given manifold can be measured by a diffusion process. Equation (A14) is the heat diffusion differential equation on a smooth manifold.

The Laplace–Beltrami-based construction described above can be replicated by means of a closely related differential operator $e^{-t\Delta_g} : L^2(M) \longrightarrow L^2(M)$ called (for $t > 0$) the *heat propagator*. Fredholm theory in this case is applied through a Green's function called the *heat kernel* $G(x, y, t)$. Similarly to the Schrödinger equation, the heat kernel is a special solution of

$$\Delta_g G(\boldsymbol{x}, \boldsymbol{y}, t) = \partial_t G(\boldsymbol{x}, \boldsymbol{y}, t), \quad \text{such that} \quad G(\boldsymbol{x}, \boldsymbol{y}, t) = e^{t\Delta_g}\delta(\boldsymbol{x} - \boldsymbol{y}).$$

The heat kernel is used to solve in integral form the diffusion equation with a source term $f(x)$ and with boundary conditions:

$$u(x, t) = \mathrm{e}^{-t\Delta_g} f(x) = \int_M G(x, y, t) f(y) \mathrm{d}\omega_g \tag{A10}$$

The heat propagator inherits from the Laplace–Beltrami operator the properties of being a self-adjoint, positive-definite and compact. Its eigenfunctions form a complete orthonormal basis of $L^2(M)$. The eigenvalue problem for the heat propagator becomes

$$\mathrm{e}^{-t\Delta_g}\varphi_i = \beta_i^t \varphi_i. \tag{A11}$$

Let us set $\lambda_i := -\ln(\beta_i)$. By construction $\mathrm{e}^{-t\Delta_g}\varphi_i$ satisfies the heat diffusion equation for all i. Substituting equation (A11), we get

$$0 = L[\mathrm{e}^{-t\Delta_g}\varphi_i] = \mathrm{e}^{-\lambda_i t}(\Delta_g\varphi_i - \lambda_i\varphi_i).$$

So φ_i are precisely the eigenfunctions of the Laplace–Beltrami operator with eigenvalues λ_i. As a consequence, the relevant properties that enable image segmentation through the Laplace–Beltrami operator (eigenfunctions define a low Dirichlet energy embedding of M in $\mathbb{R}^n$) can be directly extended to the heat propagator. In particular,

$$G(x, y, t) = \sum_i \mathrm{e}^{-\lambda_i t}\varphi_i(x)\varphi_i(y). \tag{A12}$$

For instance, the heat kernel in $\mathbb{R}^3$ is

$$G(\boldsymbol{x}, \boldsymbol{y}, t)_{\mathbb{R}^3} = \left(\frac{1}{4\pi t}\right)^{3/2} \exp\left(-\frac{\| \boldsymbol{x} - \boldsymbol{y} \|^2}{4t}\right), \quad G(\boldsymbol{x}, \boldsymbol{y}, 0) = \delta(\boldsymbol{x} - \boldsymbol{y}).$$

An argument provided in Belkin & Niyogi (2003) derives equation (A7) by discretizing the heat propagator. By substituting in equation (A10) in (A14), we get

$$\begin{aligned}\Delta_g f(x) = \Delta_g u(x, 0) &= \frac{\partial}{\partial t}\left(\int_M G(x, y, t) f(y) \mathrm{d}\omega_g\right)_{t=0} \\ &\approx -\frac{1}{t}\left(f(x) - \int_M G(x, y, t) f(y) \mathrm{d}\omega_g\right) \\ &\approx -\frac{1}{t}\left(f(x_i) - \sum_j G(x, y, t) f(x_j)\right). \end{aligned} \tag{13}$$

The overall coefficient $1/t$ does not affect the spectrum of the expression in the brackets. This argument is in line with the above discussion on the heat propagator and justifies further the diagonal term in the discrete definitions of the discrete Laplacian based on diffusion arguments.

Pointwise and spectral convergent discretizations of the heat propagator have been developed in the manifold learning theory of diffusion maps (Singer & Wu 1997; Coifman & Lafon 2006; Berry & Sauer 2016). Given the functional form of the heat kernel in $\mathbb{R}^n$, this class of discrete constructions of the Laplace–Beltrami operator pre-dominantly exploits global and local Gaussian kernels.

A4 Heat-kernel in S^2

As discussed in Section 3, we choose the adjacency weights to be the ones derived from the functional form of the integral kernel of the Laplace–Beltrami operator in S^2. In this section, we derive the functional form in equation (2).

First, we write the heat equation in spherical coordinates:

$$(\partial_t - \Delta_g)u(x, t) = L[u(x, t)] = \rho(x, t), \tag{A14}$$

with Δ_g being the Laplace–Beltrami operator in S^2, u the unknown function describing the heat diffusion, ρ the source term, x and t, respectively, the spatial and time variables. We can express Δ_g in spherical coordinates (assuming a radius $r = 1$):

$$\Delta \equiv \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}. \tag{A15}$$

Expressed in this way, we can easily recognize that the Laplace–Beltrami operator corresponds to $-\hat{L}^2$, i.e. the square of the *orbital angular momentum* operator in quantum mechanics, whose eigenfunctions are the *spherical harmonics*, $Y_{\ell m}$, and whose eigenvalues are $-\ell(\ell+1)$ ($\ell \in \mathbb{N}$). We further notice that the heat propagator operator, $\exp(-\hat{L}^2 t)$ shares the same eigenfunctions and eigenvalues as $-\hat{L}^2$. In Appendix A, we show how the integral kernel of the heat equation is related to the eigenfunctions of $\exp(-\hat{L}^2 t)$ operator (see Section A12), we can thus derive the kernel in spherical coordinates[15]:

$$\begin{aligned} G(\boldsymbol{x}, \boldsymbol{y}, t) &= \sum_{\ell=0}^{+\infty} \mathrm{e}^{-t\ell(\ell+1)} \sum_{m=-\ell}^{+\ell} Y_{\ell m}(\boldsymbol{x}) Y^*_{\ell m}(\boldsymbol{y}) \\ &= \sum_{\ell=0}^{+\infty} \frac{2\ell+1}{4\pi} \mathrm{e}^{-t\ell(\ell+1)} \mathcal{P}_\ell\left(\Theta_{\boldsymbol{xy}}\right), \end{aligned} \tag{A16}$$

where we exploit in the last equality the properties of spherical harmonics, see further details in Appendix A. In particular, the last equality is obtained by expressing the $Y_{\ell m}$ as a function of *Legendre polynomials*, $\mathcal{P}_\ell$, evaluated across the *cosine matrix*, defined as

$$\Theta_{xy} \equiv \cos\theta_{xy} = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\| \boldsymbol{x} \| \| \boldsymbol{y} \|}. \tag{A17}$$

We devote Section 3 to describe how to derive the adjacency weights from equation (A16), so that we can construct the Laplacian matrix as in equation (1) and estimate the eigenpairs to perform the spectral clustering.

A5 Laplacian eigenfunctions and the Ricci curvature

There are two conceptually distinct ways in which eigenfunctions of the Laplace–Beltrami operator and their discrete analogues are exploited for image processing.

An image is interpreted as a function on a regular domain (typically a rectangle in $\mathbb{R}^2$). This function is decomposed with respect to the special basis of eigenfunctions. This decomposition provides a useful characterization of an image, allows highlighting or filtering global components with specific Dirichlet energy i.e. spatial 'frequency'. As an elementary example the coefficient of φ_0 captures an overall intensity. This approach, similar to spectral shape analysis, has been adopted in M. Reuter & Peinecke (2005) and Rustamov (2007) and is not typically used for segmentation purposes.

An image itself is considered as smooth surface (real two-dimensional manifold) sampled in a discrete set of pixels. The geodesic distance between points on the manifold combines the distance between pixels with the 'vertical deformation' that encodes the image. Spectral geometry is a field of differential geometry that exploits the spectral structure of Laplace–Beltrami operator to describe relevant geometric and topological properties of manifolds. The spectral structure of the discrete Laplacian approximates the spectral structure of the Laplace–Beltrami operator. This is the manifold learning approach to spectral segmentation.

The application of ideas from geometry and theory of partial differential equations to image segmentation is an active field of research. Variational calculus based on a variety of energy functionals (Mumford-Shah, Chan-Vese, elastic energy with additional forcing terms and well potentials) and dynamics on different time-scales have been developed in Hu & Bertozzi (2015), Meng et al. (2017), and Boyd et al. (2018). The deep geometric nature of the spectral clustering methods is yet to be investigated. In fact, the formulation of spectral image segmentation in the smooth case means that points in plateauing regions of the manifold are close to level sets of the eigenfunctions of the Laplace–Beltrami operator in a suitable Dirichlet energy band. Relevant properties of the nodal (and more general level) sets of the eigenfunctions of Δ_g have been studied in special geometric cases (constant curvature or curvature bounded from below). Less is known on how the level sets of eigenfunctions capture the geometry of a manifold in the general case.

Equivalently, we expect that eigenfunctions of the Laplace–Beltrami operator with given Dirichlet energy are characterized by small variations on plateauing regions of the manifold. The key question is how can we characterize plateauing regions.

Our hypothesis is that these are the regions in which the Ricci curvature of the manifold is low (either positive or negative). A version of the celebrated Bochner's formula establishes a direct relation between the Ricci's curvature tensor, the Laplace–Beltrami operator and the norm of the Hessian of a smooth function φ on a Riemannian manifold M with metric g:

$$\frac{1}{2}\Delta_g |\nabla\varphi|^2 = g(\Delta_g \nabla\varphi, \nabla\varphi) + |H(\varphi)|^2 + Ric(\nabla\varphi, \nabla\varphi).$$

For eigenfunctions $\Delta_g \varphi = \lambda\varphi$, the Bochner's formula becomes

$$Ric(\nabla\varphi, \nabla\varphi) = \frac{1}{2}\Delta_g |\nabla\varphi|^2 - \lambda|\nabla\varphi|^2 - |H(\varphi)|^2.$$

For smooth surfaces (images) this equation means that the sectional curvature in the direction of the maximum variation of the eigenfunction depends directly on the magnitude of the gradient of the eigenfunction and the eigenvalue.

The above considerations suggest the existence of a deep geometric reason for the fact that spectral image segmentation is very efficient in detecting objects with complicated shapes in unsupervised set-up. Understanding spectral image segmentation in the context of spectral geometry is a promising field of future research.

APPENDIX B: CLUSTERING APPLIED ON *PLANCK*-GNILC DUST MAPS

Planck Collaboration XLVIII (2016e) released a set of maps obtained by separating the Galactic thermal dust emission from the CIB anisotropies. They implemented a tailored component-separation method, the Generalized Needlet Internal Linear Combination (GNILC) method, that adopts spatial information (e.g. the angular power spectra) to separate the Galactic dust emission and CIB anisotropies. We thus considered the temperature T_d and spectral index β_d maps of thermal dust and the uncertainties accompanying these maps to partition the sky into multiple domains accounting both for the geometrical similarities and the uncertainties.

Given the fact that we are interested to derive patches at the $\sim$ degree scales, we firstly reduce the resolution of the GNILC maps (released at `nside=2048` and 5 arcmin resolutions), to a coarser beam, $110'$ and to a lower pixel resolution, `nside=32`. This also makes the Laplacian matrix computation faster with less memory requirements.

We then performed the clustering following the Algorithm 1 separately for T_d and for β_d. In Fig. B1, we show the variances estimated from the GNILC maps in an $\alpha - \delta$ grid. We note the presence of a minimum variance vertical stripe at around $\delta \sim 0.2$ and for $\alpha \leq 0.2$. Notice that high variance regions can be identified in the upper right and lower left corners of the image related, respectively, to over and under partition regimes. Moreover, ranges of optimality both for β_d and T_d present similar variance contours. Since the optimization is performed independently for the two dust parameters,

[15] Assuming as an initial condition to the source term, ρ in (A14) to be a Dirac's delta distribution, i.e. $\rho(x, 0) = \delta(x)$.

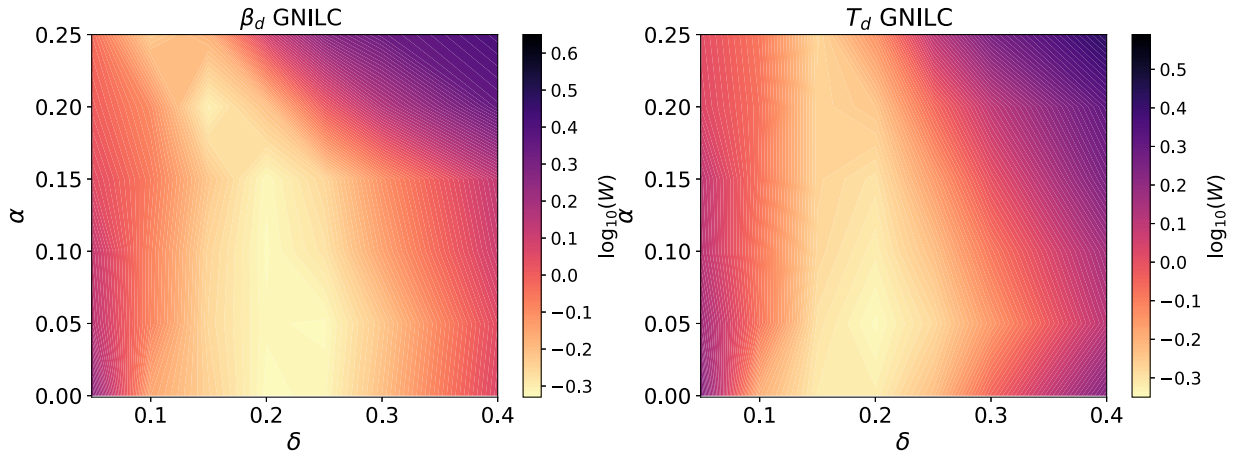

Figure B1. Variance surfaces estimated for several choices of α and δ for (left) β_d, (right) T_d. Notice that regions of optimality corresponds to local minima (lighter regions) in the surfaces.

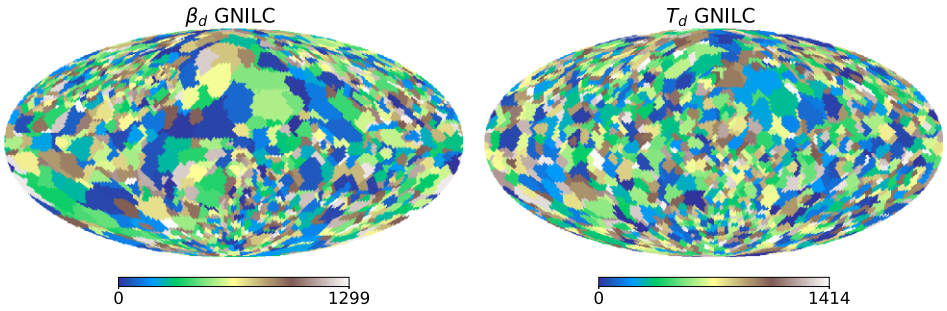

Figure B2. Cluster maps of (left) β_d and (right) T_d GNILC maps.

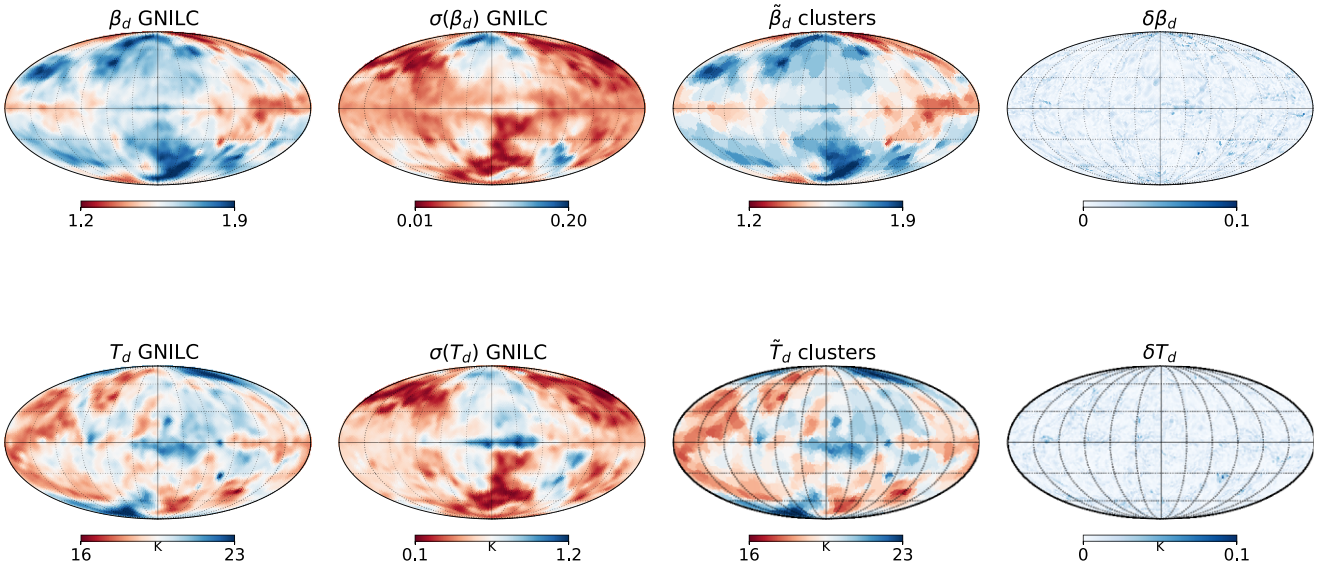

Figure B3. First and second columns: Dust parameter and uncertainty maps from *Planck*-GNILC. Third column: Dust spectral parameter maps evaluated by taking the median value of the input GNILC maps on each region defined by spectral clustering procedure. Fourth column: Relative errors estimated by taking the difference of GNILC maps with the binned ones.

we interpret this result as an indication that the methodology might be optimizing for similar Galactic scale features in both β_d and T_d maps.

We find the optimal number of clusters to be: 1300 and 1415, respectively, for β_d and T_d shown in Fig. B2, corresponding to the parameter values of $\alpha = 0.15$, $\delta = 0.20$.

Finally, we want to assess how well the optimal regions track the observed features in the GNILC maps. We thus estimate within each cluster patch, the median value of the pixels identified in a single cluster and build a *binned* map for T_d and β_d, labeled as $\tilde{T}_d$ and $\tilde{\beta}_d$. We show in Fig. B3 the maps from GNILC and the binned maps. To quantify the error of this procedure, we estimate the relative difference of the input and binned maps as in equation (12).

Relative error maps are shown in the right-most column of Fig. B3. We note that the errors are lower than the 10 per cent.

APPENDIX C: DEFINING THE OPTIMALITY RANGE

From the variances shown in Figs B1, 6 and 5, we notice that the range of optimality corresponds to a narrow range in δ but seldom involves a broader range in the α parameter.

In the top row of Fig. C1, we show several cluster configurations chosen in the range of optimality from the variance in Fig. B1 (left). We quantify the size of the cluster by estimating how many pixels belong to each cluster. The logarithm of the number of pixels per cluster is shown in the bottom row of Fig. C1. Values of $\alpha < 0.1$, lead to a very homogeneous and isotropic partition which is intuitively expected as we partition the celestial sphere given the properties of the symmetric Laplacian adjacency (equation 2). On the other hand, increasingly larger values of $\alpha > 0.2$ make the cluster morphology more and more dictated by the SNR content of the features, resulting in inhomogenuous sizes and anisotropic shapes in the vicinity of the Galactic plane as the manifold adjacency gets more and more distorted (see equation 4). We find that intermediate values for α in the range around $0.1 \div 0.2$ result in a good trade-off between these two extreme cases.

APPENDIX D: ESTIMATION OF UNCERTAINTIES FOR THE $\mathcal{N}_c$ MAP

As described in Section 4.2, we apply our clustering algorithm to partition the sky in regions of near-constant number of clouds per line of sight, as determined by the analysis of H I data in Panopoulou & Lenz (2020). A critical component of the clustering analysis is its ability to handle measurement uncertainties. The public map of $\mathcal{N}_c$ does not contain uncertainties and so we repeat steps of the original analysis to obtain estimates of the per-pixel uncertainties in the map.

We compute the uncertainty on $\mathcal{N}_c$ by taking into account the two sources of error that enter in the calculation of $\mathcal{N}_c$:

(i) The uncertainty on the number of velocity peaks identified in the H I spectrum.
(ii) The uncertainty on the column density of each cloud.

As explained in Panopoulou & Lenz (2020, Appendix B1), both sources of uncertainty are driven by the choice of the velocity kernel size (a parameter that implements smoothing in velocity space, also referred to as the bandwidth). The choice of bandwidth presents a trade-off between resolving power (ability to distinguish nearby peaks in velocity) and fidelity (avoiding spurious peak detections). Furthermore, the kernel size also alters the shape of the probability distribution, including the position of extrema and saddle points, used to define the range of each peak.

We therefore estimate the uncertainties by considering maps of $\mathcal{N}_c$ with three different choices of bandwidth: namely 3, 4, and 5 velocity channels, indicated respectively as $\mathcal{N}_c^{\rm bw3}$, $\mathcal{N}_c^{\rm bw4}$, $\mathcal{N}_c^{\rm bw5}$, with $\mathcal{N}_c \equiv \mathcal{N}_c^{\rm bw4}$ being the default value in Panopoulou & Lenz (2020). We evaluated the relative differences of $\mathcal{N}_c$ between different bandwidth runs as

$$\Delta_{rel}^{i,4} = (\mathcal{N}_c^{\rm bw\,i} - \mathcal{N}_c^{\rm bw4})/\mathcal{N}_c^{\rm bw4}. \quad \text{(D1)}$$

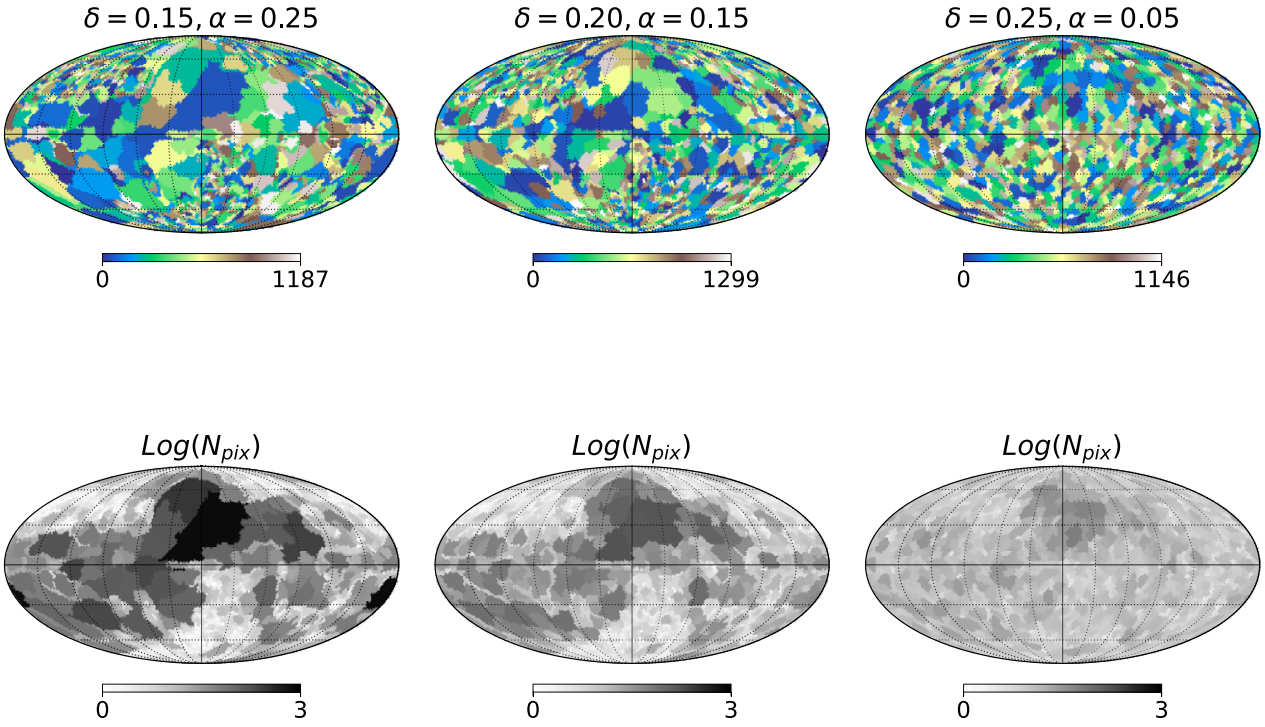

Figure C1. Top: Cluster patches for β_d GNILC maps, in regions of optimality as shown in Fig. B1, for several choices of α and δ. Bottom: Maps showing the logarithm of the number of pixels within each cluster shown in the top row.

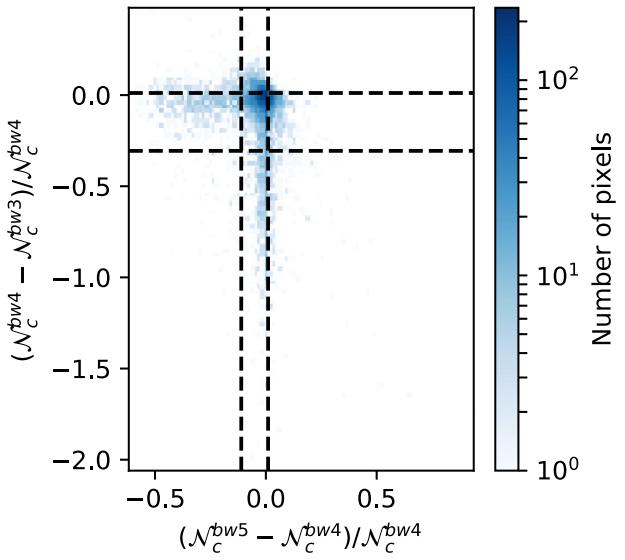

Figure D1. Distribution of asymmetric uncertainties on $\mathcal{N}_c$ at `nside=32`, using the relative difference between $\mathcal{N}_c$ measured at different values of the bandwidth parameter. Vertical axis: Difference between bandwidths of 3 and 4, normalized to the $\mathcal{N}_c^{bw4}$. Horizontal axis: Same for bandwidths 5 and 4. The dashed lines mark the 16th and 84th percentiles containing 68 per cent of the distribution of $\Delta_{rel}^{3,4}$ and $\Delta_{rel}^{5,4}$.

These differences can be used to quantify the systematic uncertainty of the cloud identification method, and hence $\mathcal{N}_c$.

Fig. D1 shows the joint distribution of $\Delta_{rel}^{4,3}$ and $\Delta_{rel}^{5,4}$ for the map at `nside=32`. The majority of pixels show relative differences of <0.1 (10 per cent). The 16 percentile and 84 percentiles of the distribution of $\Delta_{rel}^{4,3}$ are −0.3 and 0.01, respectively. For $\Delta_{rel}^{5,4}$, these values are −0.1 and 0.01. There are long tails extending out to −2 for the distribution of $\Delta_{rel}^{4,3}$ and −0.6 for that of $\Delta_{rel}^{5,4}$. We interpret these distributions as follows: by using a smaller bandwidth, a larger number of clouds is detected (due to the increased velocity resolution). The bulk of the observed differences (within 10−30 per cent relative difference) are likely related to changes in the identification of low-column density clouds. These clouds are prevalent in the H I decomposition, as discussed in Panopoulou & Lenz (2020), and contribute a low-level noise to the determination of $\mathcal{N}_c$. We will assign this low-level noise as a floor in the uncertainty on $\mathcal{N}_c$. The longer (negative) tails of the distributions are likely related to edge cases, where clouds of significant column density were unresolved at low bandwidth but become resolved with a small increase in the velocity resolution.

Because our clustering algorithm deals with Gaussian uncertainties, we need to translate these asymmetric differences into an equivalent standard deviation. We choose to err on the side of overestimating the uncertainties, by adopting the following estimate of the uncertainty on $\mathcal{N}_c$:

$$\sigma(\mathcal{N}_c) = \max\left(0.3,\ \Delta_{rel}^{3,4},\ \Delta_{rel}^{5,4}\right), \tag{D2}$$

meaning that we assign the maximum observed difference between different bandwidth runs as the (symmetric) $1-\sigma$ uncertainty in each pixel. The aforementioned low-level noise contributed by faint clouds represents the uncertainty floor that we assign (and corresponds to the absolute value of the 16th percentile of $\Delta_{rel}^{3,4}$)).

This paper has been typeset from a TEX/LATEX file prepared by the author.